Quantile Estimation of A General Single-Index Model

Efang Kong*
Technische Universiteit Eindhoven

Yingcun Xia†
National University of Singapore, Singapore

Abstract The single-index model is one of the most popular semiparametric models in Econometrics. In this paper, we define a quantile regression single-index model, which includes the single-index structure for conditional mean and for conditional variance.

Key words: Local polynomial fitting; M-regression; Strongly mixing processes; Uniform strong consistency.

1 Introduction

Regression quantiles, along with the dual methods of regression rank scores, can be considered one of the major statistical breakthroughs of the past decades. Its advantages over the other estimation methods have been well investigated. Regression quantile methods provide a much more complete statistical analysis of the stochastic relationships among variables; in addition, they are more robust against possible outliers or extreme values, and can be computed via traditional linear programming methods. Although median regression ideas go back to the 18th century and the work of Laplace, regression quantile methods were first introduced by Koenker and Bassett (1978). The linear regression quantile is very useful, but like linear regression it is not flexible enough to capture complicated relations. For quantile regression, this disadvantage

*Eurandom, Technische Universiteit Eindhoven, The Netherlands. E-mail address: kong@eurandom.tue.nl.
†Department of Statistics and Applied Probability, National University of Singapore, Singapore. http://www.stat.nus.edu.sg/~staxyc. E-mail address: staxyc@nus.edu.sg.

is even worse. As an example, consider the popular AR(1)-ARCH(1) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sigma_t z_t,\ z_t \sim \text{ IID}$$
$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2,\ \beta_0 > 0,\ \beta_1 \geq 0,$$

which cannot be fitted well by the linear quantile model.

In this paper, we focus on an important special case when the loss function is specified as

$$\rho_\tau(v) = \tau I(v > 0)v + (\tau - 1)I(v \leq 0)v, \tag{1}$$

where $0 < \tau < 1$ and $I(.)$ is the identity function, leading to the τth quantile regression, see Koenker and Bassett (1978).

In a nonparametric setting, we can state the problem as follows. Suppose Y is the response variable and $X \in R^d$ are the covariates. For loss function $\rho_\tau(.)$, we are interested in a function $m_\tau(x)$, such that

$$m_\tau(x) = \arg\min E\{\rho_\tau[Y - m(X)]\Big| X = x\} \quad \text{with respect to } m(.) \in L_1. \tag{2}$$

The function $m_\tau(x)$ is called the $\tau-$th quantile nonparametric regression function of Y on X. The application of nonparametric quantile estimation has been intensively investigated in the literature. See for example Koenker (2005) and Kong et al (2008). As in nonparametric estimation of the conditional mean function, there is the "curse of dimensionality" in estimating the typically multivariable function $m_\tau(.)$. The dimension reduction approach can thus be applied here, by considering

$$m_\tau(\theta^\top x) = \arg\min E\{\rho_\tau(Y - m(\theta^\top X))|X = x\} \quad \text{with respect to } \ \theta \in \Theta \text{ and } m(.) \in L_1, \tag{3}$$

where $\Theta = \{\theta : |\theta| = 1\}$. Ideally, we come to a single-index quantile model

$$Y = m(\theta_0^\top X) + \varepsilon, \quad E(\varphi(\varepsilon)|X) = 0,\ a.s. \tag{4}$$

where $\varphi(.)$ is the piecewise derivative function of $\rho(.)$ in (1). A typical model is the general single-index model,

$$Y = g(\theta_0^\top X, \varepsilon)$$

where ε is independent of X. Under such a model specification, it is easy to see that

$$m_\tau(x) = g_\tau(\theta^\top x) \equiv \min_v \{v : P(g(\theta_0^\top x, \varepsilon) \leq v) \geq \tau\}.$$

For the conditional heteroscadiscity model, where $g(\theta_0^\top X, \varepsilon) = g(\theta_0^\top X)\varepsilon$, we even have

$$m_\tau(x) = g(\theta_0^\top X) Q_\tau(\varepsilon)$$

where $Q_\tau(\varepsilon)$ is the $\tau-$th quantile of ε. An interesting special case for this setting is the ARCH(p) model, where $X = (y_{t-1}^2, ..., y_{t-p}^2)^\top$ and $Y = y_t$ in a time series setting.

Our main focus is the estimation of θ_0. Suppose $\{X_i, Y_i\}_{i=1}^n$ are I.I.D. observations from underlying model (4). We propose to estimate the index parameter θ_0 by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \min_{a_j, b_j} \sum_{i=1}^n \sum_{j=1}^n K(\theta^\top X_{ij}/h) \rho(Y_i - a_j - b_j \theta^\top X_{ij}), \quad X_{ij} = X_i - X_j \tag{5}$$

where $K(.)$ is a kernel function and h is a bandwidth. The minimization in (5) can be realized through iteration. First for any initial estimate $\vartheta \in \Theta$, denote by $[\hat{a}_\vartheta(x), \hat{b}_\vartheta(x)]$, the minimizer of

$$\sum_{i=1}^n K(\vartheta^\top X_{ix}/h) \rho(Y_i - a - b\vartheta^\top X_{ix}) \quad \text{with respect to } a \text{ and } b, \tag{6}$$

where $X_{ix} = X_i - x$. The estimate of θ_0 is then updated by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n \sum_{j=1}^n K(\vartheta^\top X_{ij}/h) \rho\{Y_i - \hat{a}_\vartheta(X_j) - \hat{b}_\vartheta(X_j)\theta^\top X_{ij}\}. \tag{7}$$

Repeat (6) and (7) until convergence. The true value θ_0 is thus estimated by the standardized final estimate $\hat{\theta} := \hat{\theta}/|\hat{\theta}|$.

2 Numerical studies

Again, the calculation of the above minimization problem can be decomposed into two minimization problems.

- Fixing $\theta = \vartheta$ and $w_{ij}^\vartheta = K_h(\vartheta^\top X_{ij})$, the estimation of a_j and d_j are

$$\sum_{i=1}^n \rho\{Y_i - a_j - d_j \vartheta^\top X_{ij}\} w_{ij}^\vartheta.$$

- Fixing a_j and d_j, the minimization with respect to θ can be done as follows. Again, let

$$Y_{ij}^\vartheta = Y_i (w_{ij}^\vartheta)^{1/2} - a_j (w_{ij}^\vartheta)^{1/2}, \quad X_{ij}^\vartheta = d_j X_{ij} (w_{ij}^\vartheta)^{1/2}.$$

Then the problem becomes

$$\min_{\vartheta} \sum_{i,j=1}^{n} \rho\{Y_{ij}^{\vartheta} - \theta^\top X_{ij}^{\vartheta}\}$$

Suppose the solution to the above problem is θ. Standardize it to $\theta := \theta/||\theta||$.

Set $\vartheta = \theta$ and repeat the two steps until convergence. Note that both steps are simple linear quantile regression problems and that several efficient algorithms are available, see Koenker (2005).

Example 2.1 (Single-index median regression) Consider the following model

$$y = \exp\{-5(\theta_0^\top X)^2\} + \varepsilon, \tag{8}$$

where $X \sim \Sigma_0^{1/2} X_0$ with $X_0 \sim N(0, I_5)$ and $\Sigma_0 = (0.5^{|i-j|})_{0\le i,j\le 5}$. For the noise term, we consider several distributions with both heavy tail and thin tails as well. For simplicity, we consider the median regression only. As a comparison, we also run the MAVE where a least square type estimation is used. With different sample sizes $n = 100, 200$, we carried out 100 replications. The calculation results are listed in Table 1.

Table 1: Estimation errors (and standard errors) for model (8) based on quadratic loss function and 50% quantiles

		Distribution of ε			
size	method	$0.05t(1)$	$0.1(N(0,1)^4 - 3)$	$\sqrt{5}t(5)/20$	N(0,1)/4
100	MAVE	0.3641(0.3526)	0.3530(0.3102)	0.0401(0.0182)	0.0581(0.0263)
	qMAVE	0.0902(0.1074)	0.1512(0.1957)	0.0833(0.0785)	0.1146(0.0651)
200	MAVE	0.3381(0.3389)	0.2859(0.2887)	0.0232(0.0091)	0.0373(0.0147)
	qMAVE	0.0681(0.1415)	0.0581(0.0698)	0.0402(0.0173)	0.0652(0.0272)

The MAVE method with quadratic loss function has very bad performance when the noise has heavy tail (e.g. $t(1)$) or is highly asymmetric (e.g. $N(0,1)^4$). With the absolute value loss function, the performance is much better. Even in the situation when the noise has thin tail and symmetric, qMAVE still performance reasonably well.

3 Assumptions and asymptotic properties

We adopt model (4) throughout and make the additional assumption that $\{(X_i, Y_i)\}_{i=1}^{\infty}$ are I.I.D. observations. The extension to the case of weakly dependent time series should be straightforward but complicates matters without adding anything conceptually. Furthermore, the following conditions are assumed in the proofs of Theorem 6.1.

(A1) For each $v \in \mathcal{R}$, $\rho(v)$ is absolutely continuous, i.e., there is a function $\varphi(.)$ such that $\rho(v) = \rho(0) + \int_0^v \varphi(t)dt$. The probability density function of ε_i is bounded and continuously differentiable. $E\{\varphi(\varepsilon_i)|X_i\} = 0$ almost surely and $E|\varphi(\varepsilon_i)|^{\nu_1} \le M_0 < \infty$ for some $\nu_1 > 2$.

(A2) Function $\varphi(.)$ satisfies the Lipschitz condition in (a_j, a_{j+1}), $j = 0, \cdots, m$, where $a_1 < \cdots < a_m$ are finite number of jump discontinuity points of $\varphi(.)$, $a_0 \equiv -\infty$, $a_{m+1} \equiv +\infty$ and $m < \infty$.

(A3) Kernel function $K(.)$ is symmetric density function with a compact support and satisfies $|u^j K(u) - v^j K(v)| \le C|u - v|$ for all j with $0 \le j \le 3$.

(A4) The link function $m(.)$ defined in (4) has continuous and bounded derivatives up to the third order.

(A5) The smoothing parameter h is chosen such that $nh^4 \to \infty$ and $nh^5/\log n < \infty$.

Note that (A1) and (A2) are satisfied in quantile regression with $\rho(.) = \rho_\tau(.)$ given in (1). Condition (A3) and (A4) are standard in kernel smoothing. Based on (A1) and (A2), Hong (2003) proved that there is a constant $C > 0$, such that for all small t and all x,

$$E\Big[\{\varphi(Y - t - a) - \varphi(Y - a)\}^2 | X = x\Big] \le C|t| \tag{9}$$

holds for all (a, x) in a neighborhood of $\{m(x^\top\theta_0), x\}$. Define

$$G(t; x) = E\{\rho\{Y - m(x^\top\theta_0) + t\}|X = x\}, \quad G_i(t, x) = (\partial^i/\partial t^i)G(t; x),\ i = 1, 2, 3. \tag{10}$$

Then it follows that

$$g(x) \stackrel{def}{=} G_2(0; x) \ge C > 0$$

and $G_3(t, x)$ is continuous and uniformly bounded for all $x \in \mathcal{D}$ and t near 0. For quantile regression, $g(x) = f_\varepsilon(0|x)$, where $f_\varepsilon(.|x)$ is the conditional probability density function of ε given $X = x$.

4 Initial estimator of θ_0

We use the average derivative estimation (ADE, Härdle and Stocker, 1989; Chaudhuri et al., 1997) method to obtain an initial estimate of θ_0, by observing the fact that $E[\partial m(\theta_0^\top X)/\partial X] = \theta_0 E[\partial m(\theta_0^\top X)/\partial(\theta_0^\top X)]$ and

$$\theta_0 = E[\partial m(\theta_0^\top X)/\partial X]/E[\partial m(\theta_0^\top X)/\partial(\theta_0^\top X)]. \tag{11}$$

For any $x \in R^d$ and a kernel density function $H(.) : R^d \to R^+$, denote by $[\hat{a}(x), \hat{b}(x)]$, the minimizer of the following quantity

$$\sum_{i=1}^{n} H(X_{ix}/h_0)\rho(Y_i - a - b^\top X_{ix}),$$

with respect to a and b. Observing (11), an initial estimate of θ_0 could be constructed as follows

$$\vartheta = \sum_{j=1}^{n} c(X_j)\hat{b}(X_j) \Big/ \Big| \sum_{j=1}^{n} c(X_j)\hat{b}(X_j) \Big|, \tag{12}$$

where $C(x)$ is some trimming function introduced to deal with boundary effects.

The consistency of ϑ in (12) can be proved using the results on the uniform Bahadur representation of $\hat{b}(x)$ over any compact subset $\mathcal{D}$ of the support of X. Suppose $H(.)$ is symmetric about 0 in each coordinate direction and the conditions in Proposition 3.1 and Corollary 3.3 in Kong et al (2007) are met, especially $nh_0^{d+4}/\log n < \infty$ and $nh_0^d/\log n \to \infty$. Then with probability one,

$$\hat{b}(x) = m'(\theta_0^\top x)\theta_0 + \frac{1}{nh_0^{d+1}\{fg\}(x)} \sum_{i=1}^{n} H(X_{ix}/h_0)\varphi(\varepsilon_i)X_{ix}/h_0 + O\Big\{h_0^{-1}\Big(\frac{\log n}{nh_0^d}\Big)^{3/4}\Big\} \tag{13}$$

uniformly in $x \in \mathcal{D}$, where $\{fg\}(x) = f(x)g(x)$ with $f(.)$ the density function of X and $g(x) > 0$ some deterministic function. This in turn implies that with probability one,

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} c(X_j)\hat{b}(X_j) \quad &= \quad m'(\theta_0^\top x)\theta_0 + \frac{1}{n^2 h_0^{d+1}} \sum_{i,j=1}^{n} c(X_j)\{fg\}^{-1}(X_j)H(X_{ij}/h_0)\varphi(\varepsilon_i)X_{ij}/h_0 \\ &\quad + O\Big\{h_0^{-1}\Big(\frac{\log n}{nh_0^d}\Big)^{3/4}\Big\}. \end{aligned}$$

Using results in Masry (1996), we know that with probability 1,

$$\frac{1}{nh_0^d}\sum_{i=1}^{n} H(X_{ix}/h_0)\varphi(\varepsilon_i)\frac{X_{ix}}{h_0} = O\{(nh_0^d/\log n)^{-1/2}\}$$

uniformly in $x \in \mathcal{D}$, whence

$$\frac{1}{n^2 h_0^{d+1} 2}\sum_{i,j=1}^{n} c(X_j)\{fg\}^{-1}(X_j)H(X_{ij}/h_0)\varphi(\varepsilon_i)\frac{X_{ij}}{h_0} = O\{h_0^{-1}(nh_0^d/\log n)^{-1/2}\}$$

almost surely. Therefore, concerning the initial estimator ϑ in (12), we have

$$\delta_\vartheta \equiv \theta_0 - \vartheta = O\{h_0^{-1}(nh_0^d/\log n)^{-1/2}\} \tag{14}$$

almost surely. Consequently from now on, we focus on parametric space $\Theta_n \equiv \{\vartheta : |\delta_\vartheta| < Ch(nh_0^{d+2}/\log n)^{-1/2}\}$ for some constant $C > 0$.

5 Asymptotics of $\hat{a}_\vartheta(x)$ and $\hat{b}_\vartheta(x)$

For any $\vartheta \in \Theta_n$, denote by $f_\vartheta(x)$ and $F_\vartheta(x)$,the probability density function and distribution function of $\vartheta^\top X$ at $\vartheta^\top x$ respectively, and for any $v \in R$ and $x \in \mathcal{D} \subset R^d$, define

$$\begin{aligned}
&m_\vartheta(v) = \arg\min_a E\{\rho(Y-a)|X^\top\vartheta = v\},\\
&G_\vartheta(t,x) = E\{\rho(Y - m_\vartheta(\vartheta^\top x) + t)|\vartheta^\top X = \vartheta^\top x\},\\
&G^i_\vartheta(t,x) = (\partial^i/\partial t^i)G_\vartheta(t,x),\ i=1,2; \quad g_\vartheta(x) = G^2_\vartheta(m_\vartheta(x),x)
\end{aligned}$$

Apparently $g_{\theta_0}(x) \equiv g(x)$. We assume that for any ϑ in a neighborhood of θ_0, $G^2_\vartheta(t,x)$ is continuous and uniformly bounded in the neighborhood of $(m_\vartheta(x),x)$ and there exists some $\delta > 0$ such that $g_\vartheta(x) > \delta$ for ϑ near enough θ_0 and $x \in \mathcal{D}$.

With initial estimate ϑ, let $[\hat{a}_j, \hat{b}_j] \equiv [\hat{a}_\vartheta(X_j), \hat{b}_\vartheta(X_j)]$ be the solution to (6) with x specified as X_j. If the smoothing parameter h is chosen such that $nh/\log n \to \infty$ and $nh^5/\log n < \infty$, using the results on uniform Bahadur representation in Kong et al (2007), we have

$$\begin{aligned}
\hat{a}_j - m_\vartheta(X_j) &= \frac{1}{nh}\{g.f\}^{-1}_\vartheta(X_j)\sum_{i=1}^n K^\vartheta_{ij}\varphi(Y^*_{ij}) + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\}, \qquad (15)\\
h\{\hat{b}_j - m'_\vartheta(X_j)\} &= \frac{1}{nh}\{g.f\}^{-1}_\vartheta(X_j)\sum_{i=1}^n K^\vartheta_{ij}\varphi(Y^*_{ij})X^\top_{ij}\vartheta/h + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\},
\end{aligned}$$

uniformly in $X_j \in \mathcal{D}$, where $K^\vartheta_{ij} = K(X^\top_{ij}\vartheta/h)$, $Y^*_{ij} = Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X^\top_{ij}\vartheta$ and $\{g.f\}_\vartheta(.) = g_\vartheta(.)f_\vartheta(.)$. Note that $m_\vartheta(X_j) \overset{def}{=} m_\vartheta(X_j^\top\vartheta)$ and $m'_\vartheta(X_j) \overset{def}{=} m'_\vartheta(X_j^\top\vartheta)$.

Combined with Lemma 6.5 and Lemma 6.6 in the Appendix, further to (15), we have

$$\begin{aligned}
\hat{a}_j - a_j &= \frac{1}{2}m''(X_j^\top\theta_0)_\vartheta(X_j)h^2 + b_j\delta^\top_\vartheta\{(\nu/\mu)_\vartheta(X_j) - X_j\}\\
&\quad + (nh)^{-1}\{gf\}^{-1}_\vartheta(X_j)\sum_{i=1}^n\varphi_{ij} + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4} + h^4 + h\delta_\vartheta\Big\}, \qquad (16)\\
\hat{b}_j - b_j &= h^2\Big[\frac{1}{2}m''(X_j^\top\theta_0)\{(f\mu)'/(fg)\}_\vartheta(X_j) + \frac{1}{6}m^{(3)}(X_j^\top\theta_0)\{(f\mu)/(fg)\}_\vartheta(X_j)\Big]\\
&\quad + b_j\delta^\top_\vartheta\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + (nh^2)^{-1}\{gf\}^{-1}_\vartheta(X_j)\sum_{i=1}^n\tilde{\varphi}_{ij}\\
&\quad + O\Big\{h^4 + h^2\delta_\vartheta + \Big(\frac{\log n}{nh}\Big)^{3/4}/h\Big\}
\end{aligned}$$

uniformly in j with $X_j \in \mathcal{D}$, where $(\nu/\mu)_\vartheta(X_j) \equiv \nu_\vartheta(X_j^\top\vartheta)/\mu_\vartheta(X_j^\top\vartheta)$,

$$\mu_\vartheta(v) = E[g(X)|X^\top\vartheta = v], \quad \nu_\vartheta(v) = E[g(X)X|X^\top\vartheta = v]. \qquad (17)$$

and φ_{ij} and $\tilde{\varphi}_{ij}$ are zero-mean I.I.D. random variables defined as

$$\begin{aligned} \varphi_{ij} &= K_{ij}^{\vartheta}\varphi(Y_{ij}^*) - E[K_{ij}^{\vartheta}\varphi(Y_{ij}^*)], \\ \tilde{\varphi}_{ij} &= K_{ij}^{\vartheta}\varphi(Y_{ij}^*)X_{ij}^{\top}\vartheta/h - E[K_{ij}^{\vartheta}\varphi(Y_{ij}^*)X_{ij}^{\top}\vartheta/h]. \end{aligned} \tag{18}$$

Note that (16) focuses on the almost sure property of $[\hat{a}_j, \hat{b}_j]$. Welsh (1996) studied their the asymptotic bias and variance, i.e.

$$\begin{aligned} &E\{\hat{a}(x)\} = m_{\vartheta}(\vartheta^{\top}x) + O(h^2), \quad E\{\hat{b}(x)\} = m_{\vartheta}'(\vartheta^{\top}x) + O(h^2), \\ &\mathrm{Var}\{\hat{a}(x)\} = O(n^{-1}h^{-3}), \quad \mathrm{Var}\{\hat{b}(x)\} = O(n^{-1}h^{-3}), \end{aligned} \tag{19}$$

and the $O(.)$s are uniformly in x in any compact subset of the support of X.

6 Asymptotics of $\hat{\theta}$

For the previously obtained ϑ, $\hat{a}_j,\ \hat{b}_j,\ j = 1, \cdots, n$, suppose $\hat{\theta}$ minimizes $\tilde{\Phi}_n(\theta)$, where

$$\sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}^{\vartheta}\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^{\top}X_{ij}) + \frac{n^2h}{2}(\theta-\vartheta)^{\top}\vartheta\vartheta^{\top}(\theta-\vartheta).$$

Apparently, $\hat{\theta}$ also minimizes

$$\begin{aligned} \tilde{\Phi}_n(\theta) &= \Phi_n(\theta) + n^2h\{\frac{1}{2}(\theta-\theta_0)^{\top}\vartheta\vartheta^{\top}(\theta-\theta_0) + (\theta_0-\vartheta)^{\top}\vartheta\vartheta^{\top}(\theta-\theta_0)\} \\ \Phi_n(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}^{\vartheta}\{\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^{\top}X_{ij}) - \rho(Y_{ij})\}, \end{aligned} \tag{20}$$

where $Y_{ij} \equiv Y_i - \hat{a}_j - \hat{b}_jX_{ij}^{\top}\theta_0$. Let $a_{n\vartheta} = \max\{(n\log\log n)^{-1/2}, |\delta_{\vartheta}|\}$. As $|\vartheta - \theta_0| = O(a_{n\vartheta})$, $\vartheta\vartheta^{\top} = \theta_0\theta_0^{\top} + O(a_{n\vartheta})$, whence for any θ with $\delta_{\theta} \overset{def}{=} \theta_0 - \theta = O(a_{n\vartheta})$, we have

$$\tilde{\Phi}_n(\theta) = \Phi_n(\theta) + n^2h\{\frac{1}{2}\delta_{\theta}^{\top}\theta_0\theta_0^{\top}\delta_{\theta} - \delta_{\vartheta}^{\top}\theta_0\theta_0^{\top}\delta_{\theta}\} + o(n^2ha_{n\vartheta}^2).$$

Write $\Phi_n(\theta) = E[\Phi_n(\theta)] + \delta_{\theta}^{\top}\{R_{n1}(\theta) - ER_{n1}(\theta)\} + R_{n2}(\theta) - ER_{n2}(\theta)$, where

$$R_{n1} = \sum_{i,j} K_{ij}^{\vartheta}\varphi(Y_{ij})\hat{b}_jX_{ij},\ R_{n2}(\theta) = \sum_{i,j} K_{ij}^{\vartheta}\Big[\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^{\top}X_{ij}) - \rho(Y_{ij}) - \delta_{\theta}^{\top}\varphi(Y_{ij})\hat{b}_jX_{ij}\Big].$$

Applying the results on $E(\Phi_n(\theta))$ in Lemma 6.11, we have

$$\Phi_n(\theta) = \delta_{\theta}^{\top}R_{n1} + \frac{1}{2}\delta_{\theta}^{\top}G_{n\vartheta}\delta_{\theta}\{1+o(1)\} + R_{n2}(\theta) - ER_{n2}(\theta), \tag{21}$$

where

$$G_{n\vartheta} = \sum_{i,j} E[K_{ij}^{\vartheta} g(X_i)\hat{b}_j^2 X_{ij} X_{ij}^{\top}] = n^2 h S_2\{1 + O(\delta_{\vartheta})\},$$

$$S_2 = \int \{m'(X^{\top}\theta_0)\}^2 \omega_{\theta_0}(X) f_{\theta_0}(X) dX,$$

and $\omega_{\vartheta}(x) = E\{g_{\vartheta}(X)(X - x)(X - x)^{\top} | X^{\top}\vartheta = x^{\top}\vartheta\}$. Consequently,

$$\tilde{\Phi}_n(\theta) = (R_{n1} - \theta_0\theta_0^{\top})\delta_{\theta} + \frac{1}{2}\delta_{\theta}^{\top}(G_{n\vartheta} + n^2 h\theta_0\theta_0^{\top})\delta_{\theta}\{1 + o(1)\} + R_{n2}(\theta) - ER_{n2}(\theta).$$

Our main result is as follows

Theorem 6.1 *Suppose (A1)-(A4) hold. With $\nu_{\vartheta}(.)$ and $\mu_{\vartheta}(.)$ as defined in (17), we have*

$$\begin{aligned}
\hat{\theta} - \theta_0 &= (S_2 + \theta_0\theta_0^{\top})^{-1}\frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) \\
&\quad -(S_2 + \theta_0\theta_0^{\top})^{-1}(\Omega_{n\vartheta} + \theta_0\theta_0^{\top})\delta_{\vartheta} + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2}) \\
&= (S_2 + \theta_0\theta_0^{\top})^{-1}\frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) \\
&\quad -(S_2 + \theta_0\theta_0^{\top})^{-1}(\Omega_0 + \theta_0\theta_0^{\top})\delta_{\vartheta} + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2}) \qquad (22)
\end{aligned}$$

almost surely, where $\varpi_{\theta}(x) = E(X|X^{\top}\theta = x^{\top}\theta) - x$, $\alpha_n = o(1)$ uniformly in ϑ and

$$\Omega_{n\vartheta} \stackrel{def}{=} \frac{1}{n}\sum_j b_j^2 \mu_{\vartheta}(X_j)\{(\nu/\mu)_{\vartheta}(X_j) - X_j\} \times \{(\nu/\mu)_{\vartheta}(X_j) - X_j\}^{\top}$$

$$\Omega_0 = E[\{m'(X^{\top}\theta_0)\}^2 \mu_{\theta_0}(X)\{(\nu/\mu)_{\theta_0}(X) - X\}\{(\nu/\mu)_{\theta_0}(X) - X\}^{\top}]$$

Remark 6.2 In Lemma 6.16, we prove that if $\delta_{\vartheta} \neq 0$,

$$0 < |(S_2 + \theta_0\theta_0^{\top})^{-1}(\Omega_0 + \theta_0\theta_0^{\top})\delta_{\vartheta}|/|\delta_{\vartheta}| < 1. \qquad (23)$$

This implies that the effect on $\hat{\theta} - \theta_0$ of the initial estimate error $\vartheta - \theta_0$ decreases geometrically.

Remark 6.3 Theorem 6.1 is proved under the assumption that $\{(X_i, Y_i)\}_{i=1}^{\infty}$ are I.I.D. observations. It is possible, however, to extend this result for time series observations provided that the time dependency (usually measured by mixing coefficient) are weak enough. For example, the stationary $\beta-$ mixing processes, which satisfies

$$\beta(k) = \sup_{A\in\mathcal{F}_{-\infty}^{a}, B\in\mathcal{F}_{a+k}^{\infty}} |P(B) - P(B|A)| \to 0, \quad \text{as } k \to \infty,$$

where $\mathcal{F}_a^b$ is the $\sigma-$algebra generated by $\{(X_i, Y_i)\}_{i=a}^{b}$.

Lemma 6.4 *Under conditions in Theorem 6.1, we have*

$$(n^2h)^{-1}R_{n1} = \frac{1}{n}\sum_i \varphi(\varepsilon_i)b_i\{\varpi f\}_{\theta_0}(X_i) - \Omega_{n\vartheta}\delta_\vartheta + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2})\ a.s. \tag{24}$$

Proof of Theorem 6.1. Based on (24), it suffices to prove that

$$\hat\theta - \theta_0 = \{n^2h(S_2 + \theta_0\theta_0^\top)\}^{-1}(R_{n1} - n^2h\theta_0\theta_0^\top\delta_\vartheta)\ a.s. \tag{25}$$

As the first step to prove (25), we show in Lemma 6.13 and Lemma 6.14 that for each fixed θ,

$$(n^2ha_{n\vartheta}^2)^{-1}[R_{n2}(\theta) - ER_{n2}(\theta)] = o(1)\ a.s. \tag{26}$$

This together with (21) and the fact that $G_{n\vartheta} = n^2hS_2\{1 + O(\delta_\vartheta)\}$ imply that for any fixed θ,

$$(n^2ha_{n\vartheta}^2)^{-1}[\tilde\Phi_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta] \to 0\ a.s.$$

As both $\tilde\Phi_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta)$ and $\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta$ are convex in θ, it follows from Lemma 6.7 that for any compact set $\Theta_{n\theta} \subset \Theta_n$(convex open set),

$$\sup_{\theta\in\Theta_{n\theta}}(n^2ha_{n\vartheta}^2)^{-1}|\tilde\Phi_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta| \to 0\ a.s. \tag{27}$$

Let $\eta_n = \{n^2h(S_2 + \theta_0\theta_0^\top)\}^{-1}(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta)$. Now we are ready to prove the equivalent of (25), i.e. with probability 1, for any $\delta > 0$, $|\hat\theta - \theta_0 - \eta_n|/a_{n\vartheta} \le \delta$ for large n.

First note that as $\theta_0 + \eta_n$ is bounded with probability 1, Θ_n can be chosen to contain B_n^δ, a closed ball with center $\theta_0 + \eta_n$ and radius $a_{n\vartheta}\delta$. Replace $\Theta_{n\theta}$ in (27) by B_n^δ, we have

$$\Delta_n \equiv \sup_{\theta\in B_n^\delta}(n^2ha_{n\vartheta}^2)^{-1}|\tilde\Phi_n(\theta) - \delta_\theta^\top(R_{n1} - \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta| = o(1)\ a.s. \tag{28}$$

Now consider the behavior of $\tilde\Phi_n(\theta)$ outside B_n^δ. Suppose $\theta = \theta_0 + \eta_n + a_{n\vartheta}\beta\nu$, for some $\beta > \delta$ and ν a unit vector. Define θ^* as the boundary point of B_n^δ that lies on the line segment from $\theta_0 + \eta_n$ to θ, i.e. $\theta^* = \theta_0 + \eta_n + a_{n\vartheta}\delta\nu$. Convexity of $\Phi_n(\theta)$ and the definition of Δ_n imply

$$\begin{aligned}
\frac{\delta}{\beta}\tilde\Phi_n(\theta) + (1 - \frac{\delta}{\beta})\tilde\Phi_n(\theta_0 + \eta_n) \quad &\ge \quad \tilde\Phi_n(\theta^*) \\
&\ge \quad \frac{1}{2}n^2h\delta^2a_{n\vartheta}^2\nu^\top(S_2 + \theta_0\theta_0^\top)\nu \\
&\quad\ -\frac{1}{2}(n^2h)^{-1}R_{n1}^\top(S_2 + \theta_0\theta_0^\top)^{-1}R_{n1} - n^2ha_{n\vartheta}^2\Delta_n \\
&\ge \quad \frac{1}{2}n^2h\delta^2a_{n\vartheta}^2\nu^\top(S_2 + \theta_0\theta_0^\top)\nu + \tilde\Phi_n(\theta_0 + \eta_n) - 2n^2ha_{n\vartheta}^2\Delta_n.
\end{aligned}$$

It follows that

$$\inf_{|\theta-\theta_0-\eta_n|>\delta a_{n\vartheta}} \tilde{\Phi}_n(\theta) \geq \tilde{\Phi}_n(\theta_0+\eta_n) + \frac{\beta}{\delta} n^2 h a_{n\vartheta}^2 [\frac{1}{2}\delta^2 \nu^\top (S_2+\theta_0\theta_0^\top)\nu - 2\Delta_n].$$

As $S_2+\theta_0\theta_0^\top$ is positive definite, then according to (28), with probability 1, $\delta^2\nu^\top S_2 \nu > 4\Delta_n$ for large enough n. This implies that for any $\delta > 0$ and for large enough n, the minimum of $\tilde{\Phi}_n(\theta)$ must occur within B_n^δ. This implies (25). ■

Appendix

Proof of Lemma 6.4. Write

$$R_{n1}(\theta) = \sum_{i,j} K_{ij}^\vartheta \varphi(\varepsilon_i) b_j X_{ij} + \sum_{i,j} K_{ij}^\vartheta \varphi(\varepsilon_i)(\hat{b}_j - b_j) X_{ij} + \sum_{i,j} K_{ij}^\vartheta \hat{b}_j X_{ij}\{\varphi(Y_{ij}) - \varphi(\varepsilon_i)\},$$

where E_j denotes expectation taken w.r.t X_j for given X_i. We will show that

$$\frac{1}{n^2 h}\sum_{i,j} K_{ij}^\vartheta \varphi(\varepsilon_i) b_j X_{ij} = \frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) + O\{(\log\log n/n)^{1/2}(h^2+\delta_\vartheta)\}, \tag{29}$$

which together with Lemma 6.12 lead to (24).

First note that

$$\begin{aligned} E_j[K_{ij}^\vartheta b_j X_{ij}/h] = b_i\{\varpi f\}_\vartheta(X_i) - \delta_\vartheta m''(X_i^\top\theta_0)\{\Sigma f\}_{\theta_0}(X_i) \\ +h^2 b_i\{\varpi f\}_{\theta_0}''(X_i) + O(|\delta_\vartheta|^2 + h^4), \end{aligned}$$

This together with Lemma 7.8 in Xia and Tong (2006), we have

$$\frac{1}{n^2 h}\sum_{i,j} K_{ij}^\vartheta \varphi(\varepsilon_i) b_j X_{ij} = \frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) + O\{(\log\log n/n)^{1/2}(h^2+\delta_\vartheta)\},$$

from which follows (29), as $\{\varpi f\}_\vartheta(.)$ is lipschitz continuous in ϑ. ■

Lemma 6.5

$$\begin{aligned} m_\vartheta(X_j) - a_j &= b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\} + o(|\delta_\vartheta|), & (30)\\ m_\vartheta'(X_j) - b_j &= b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + o(|\delta_\vartheta|), & (31) \end{aligned}$$

Proof It follows from the property of conditional expectation that

$$\begin{aligned} E\{\rho(Y-a)|X^\top\vartheta = x^\top\vartheta\} &= E[E\{\rho(Y-a)|X\}|X^\top\vartheta = x^\top\vartheta] \\ &= E[G\{m(\theta_0^\top X) - a; X\}|X^\top\vartheta = x^\top\vartheta]. \end{aligned}$$

Using the differentiability of $G(t;X)$ in t, we have

$$G\{m(\theta_0^\top X)-a;X\}=G(0;X)+g(X)(m(\theta_0^\top X)-a)^2/2+O\{(m(\theta_0^\top X)-a)^3\}.$$

If $X^\top\vartheta=x^\top\vartheta$ and $\delta_\vartheta=o(1)$, $m(\theta_0^\top X)-m(\theta_0^\top x)=O\{\theta_0^\top(X-x)\}=O\{\delta_\theta^\top(X-x)\}=o(1)$. Therefore for every a near $m(\theta_0^\top X)$ (whence $m(\theta_0^\top x)$ [WHY]),

$$\begin{aligned}&E[G\{m(\theta_0^\top X)-a;X\}|X^\top\vartheta=x^\top\vartheta]-E[G(0;X)|X^\top\vartheta=x^\top\vartheta]\\&\qquad\qquad\qquad\to\frac{1}{2}E[g(X)(m(\theta_0^\top X)-a)^2|X^\top\vartheta=x^\top\vartheta].\end{aligned}$$

As $\rho(.)$ is convex, we can argue this convergence is in fact uniform over all a near $m(\theta_0^\top X)$, which implies that the minima of $E[G\{m(\theta_0^\top X)-a;X\}|X^\top\vartheta=x^\top\vartheta]$ is also approximately the minima of $E[g(X)(m(\theta_0^\top X)-a)^2|X^\top\vartheta=x^\top\vartheta]$. We have

$$\begin{aligned}m(\theta_0^\top X)=m(\theta_0^\top x)+m'(\theta_0^\top x)\theta_0^\top(X-x)+C\{\theta_0^\top(X-x)\}^2,&\\ E[g(X)(m(\theta_0^\top X)-a)^2|X^\top\vartheta=x^\top\vartheta]\;=\;2m'(\theta_0^\top x)\{m(\theta_0^\top x)-a\}\delta_\vartheta^\top\{\nu_\vartheta(x^\top\vartheta)-x\mu_\vartheta(x^\top\vartheta)\}&\\ +\{m(\theta_0^\top x)-a\}^2\mu_\vartheta(x^\top\vartheta)+O(|\delta_\vartheta|^2).&\qquad(32)\end{aligned}$$

Take derivative with respect to a and (30) follows.

To prove (31), for any $t\to 0$, mimicking (32),

$$\begin{aligned}&E[g(X)\{m(\theta_0^\top X)-a\}^2|X^\top\vartheta=x^\top\vartheta+t]\\=\;&2m'(\theta_0^\top x)\{m(\theta_0^\top x)-a\}E[g(X)\{t+\delta_\vartheta^\top(X-x)\}|X^\top\vartheta=x^\top\vartheta+t]\\&+\{a-m(\theta_0^\top x)\}^2\mu_\vartheta(x^\top\vartheta+t)+O(|\delta_\vartheta|^2)\\=\;&\{a-m(\theta_0^\top x)\}^2\mu_\vartheta(x^\top\vartheta+t)+2tm'(\theta_0^\top x)\{m(\theta_0^\top x)-a\}\mu_\vartheta(x^\top\vartheta+t)+O(t^2|\delta_\vartheta|^2)\\&+2m'(\theta_0^\top x)\{m(\theta_0^\top x)-a\}\delta_\vartheta^\top\{\nu_\vartheta(x^\top\vartheta+t)-x\mu_\vartheta(x^\top\vartheta+t)\}.\end{aligned}$$

Again take derivative with respect to a and by the definition of $m_\vartheta(.)$, we have

$$m_\vartheta(\vartheta^\top x+t)\approx m(\theta_0^\top x)+tm'(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(x^\top\vartheta+t)-x\},$$

Recall that from (30), $m_\vartheta(\vartheta^\top x)\approx m(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(x^\top\vartheta)-x\}+O(|\delta_\vartheta|^2)$. Subtract this from the equation above and suppose the first order derivative of $\mu_\vartheta(.)$ and $\nu_\vartheta(.)$ are both Lipschitz continuous, we have

$$\begin{aligned}&m_\vartheta(\vartheta^\top x+t)-m_\vartheta(\vartheta^\top x)\\\approx\;&tm'(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(x^\top\vartheta+t)-(\nu/\mu)_\vartheta(x^\top\vartheta)\}\\=\;&tm'(\theta_0^\top x)+tm'(\theta_0^\top x)\delta_\vartheta^\top\{(\mu\nu'-\mu'\nu)/u^2\}_\vartheta(x^\top\vartheta)+O(t^2).\end{aligned}$$

Divide this over t and let $t \to 0$, we will have (31). ■

Lemma 6.6 $E_i K_{ij}^{\vartheta} \varphi(Y_{ij}^*) = \frac{1}{2} m''(X_j^\top \theta_0)(fg)_\vartheta(X_j) h^3 + O(h^4) + o(h\delta_\vartheta)$,

$$
\begin{aligned}
E_i K_{ij}^{\vartheta} \varphi(Y_{ij}^*) X_{ij}^\top \vartheta &= h^4 \Big\{ \frac{1}{2} m''(X_j^\top \theta_0)(f\mu)'_\vartheta(X_j) \\
&\quad + \frac{1}{6} m^{(3)}(X_j^\top \theta_0)(f\mu)_\vartheta(X_j) \Big\} + O(h^4 \delta_\vartheta + h^6). \qquad (33)
\end{aligned}
$$

Proof Based on (30) and (31), we have

$$
\begin{aligned}
& m(X_i^\top \theta_0) - m_\vartheta(X_j) - m'_\vartheta(X_j) X_{ij}^\top \vartheta \\
= \; & m(X_i^\top \theta_0) - m(X_j^\top \theta_0) - b_j \delta_\vartheta^\top \{(\nu/\mu)_\vartheta(X_j) - X_j\} \\
& - \{b_j + b_j \delta_\vartheta^\top \{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j)\} X_{ij}^\top \vartheta + o(|\delta_\vartheta|) \\
= \; & b_j X_{ij}^\top \delta_\vartheta + \frac{1}{2} m''(X_j^\top \theta_0)(\theta_0^\top X_{ij})^2 + \frac{1}{6} m^{(3)}(X_j^\top \theta_0)(\theta_0^\top X_{ij})^3 \\
& - b_j \delta_\vartheta^\top \{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) X_{ij}^\top \vartheta - b_j \delta_\vartheta^\top \{(\nu/\mu)_\vartheta(X_j) - X_j\} \\
& + o(|\delta_\vartheta|) + O\{(X_{ij}^\top \vartheta)^4 + \delta_\vartheta\}.
\end{aligned}
$$

As $m(X_i^\top \theta_0) - m_\vartheta(X_j) - m'_\vartheta(X_j) X_{ij}^\top \vartheta = o(1)$, by the continuity of $G_1(t; X)$ in t, we have

$$
\begin{aligned}
& E[\varphi\{Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j) X_{ij}^\top \vartheta\} | X_i] \\
= \; & G_1\{m(X_i^\top \theta_0) - m_\vartheta(X_j) - m'_\vartheta(X_j) X_{ij}^\top \vartheta; X_i\} \\
= \; & b_j \delta_\vartheta^\top g(X_i) X_{ij} - b_j \delta_\vartheta^\top \{(\nu/\mu)_\vartheta(X_j) - X_j\} g(X_i) - b_j \delta_\vartheta^\top \{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) g(X_i) X_{ij}^\top \vartheta \\
& + \frac{1}{2} m''(X_j^\top \theta_0) g(X_i)(\theta_0^\top X_{ij})^2 + \frac{1}{6} m^{(3)}(X_j^\top \theta_0) g(X_i)(\theta_0^\top X_{ij})^3 + o(|\delta_\vartheta|) + O((X_{ij}^\top \vartheta)^4), \quad (34)
\end{aligned}
$$

and thus

$$
E_i[K_{ij}^{\vartheta} \varphi\{Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j) X_{ij}^\top \vartheta\}] = \frac{1}{2} m''(X_j^\top \theta_0)(gf)_\vartheta(X_j) h^3 + o(h|\delta_\vartheta|) + O(h^4).
$$

Similarly (33) follows from (34) and the following facts

$$
\begin{aligned}
E[g(X_i)X_{ij}|X_i^\top\vartheta = X_j^\top\vartheta + hu] &= \nu_\vartheta(X_j^\top\vartheta + hu) - X_j\mu_\vartheta(X_j^\top\vartheta + hu) \\
&= \nu_\vartheta(X_j^\top\vartheta) + hu\nu_\vartheta'(X_j^\top\vartheta) - X_j\mu_\vartheta(X_j^\top\vartheta) \\
&\quad -huX_j\mu_\vartheta'(X_j^\top\vartheta) + O(h^2), \\
E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu] &= \mu_\vartheta(X_j^\top\vartheta) + hu\mu_\vartheta'(X_j^\top\vartheta) + O(h^2), \\
\int K(u)E[g(X_i)X_{ij}|X_i^\top\vartheta = X_j^\top\vartheta + hu]hudu &= h^2\{(f\nu')_\vartheta(X_j^\top\vartheta) - X_j(f\mu')_\vartheta(X_j^\top\vartheta)\} \\
&\quad +h^2\{(f'\nu)_\vartheta(X_j^\top\vartheta) - X_j(f'\mu)_\vartheta(X_j^\top\vartheta)\} + O(h^4), \\
\int K(u)E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu]hudu &= h^2(\mu'f + \mu f')_\vartheta(X_j^\top\vartheta) + O(h^4), \\
\int K(u)E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu]h^2u^2du &= h^2(\mu f)_\vartheta(X_j^\top\vartheta) + O(h^4). \qquad \blacksquare
\end{aligned}
$$

Lemma 6.7 *Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a sequence of random convex functions defined on a convex, open subset Θ of R^d. Suppose $\lambda(\theta)$ is a real valued function on Θ such that $\lambda_n(\theta)$ tends to $\lambda(\theta)$ for each θ almost surely, Then for each compact set K of Θ, with probability* 1,

$$\sup_{\theta\in K} |\lambda_n(\theta) - \lambda(\theta)| \to 0.$$

Proof The condition can be restated as follows: for any fixed $\theta \in \Theta$, there exists some $\Omega_\theta \subseteq \Omega$, such that $P(\Omega_\theta) = 1$ and

$$\lambda_n(\omega, \theta) - \lambda(\theta) \to 0, \text{ for any } \omega \in \Omega_\theta.$$

The conclusion can be restated that for each compact set K of Θ, there exists some $\Omega_0 \subseteq \Omega$, such that

$$P(\Omega_0) = 1 \quad \text{and } \sup_{\theta\in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \to 0, \text{ for any } \omega \in \Omega_0.$$

For such uniformity of the convergence, it is enough to consider the case where K is a cube with edges parallel to the coordinate directions $e_l, \cdots, e_d$. Every compact subset of Θ can be covered by finitely many such cubes. Let $\Im_0 \equiv K$ and $K^{+\delta_0}$ be the larger cube constructed by adding an extra layer of cubes with sides δ_0 to K. Suppose $\delta_0 > 0$ is small enough such that $K^{+\delta_0} \subset \Theta$. Define $\mho_0$ for the finite set of all vertices of all the cubes that make up $K^{+\delta^0}$.

Now for $k = 1, 2, \cdots$, let $\epsilon_k = k^{-1}$. As convexity implies continuity, there is a $0 < \delta^k < \delta^{k-1}$ such that $\lambda(.)$ varies by less than $\epsilon_k/(d+1)$ over each cube of side $3\delta^k$ that intersects K. Partition

each cube in $\Im_{k-1}$ into a union of cubes with side at most δ^k and denote by $\Im_k$ the resulted union of cubes. Then expand K to a larger cube $K^{+\delta^k}$ by adding an extra layer of these δ^k-cubes around each face. As $\delta^k < \delta^{k-1}$, $K^{+\delta^k} \subset K^{+\delta^{k-1}}$ is still within Θ. Define

$$\begin{aligned}
\mho_k &= \{ \text{ vertices of all the } \delta^k - \text{cubes that make up } K^{+\delta^k}\} \bigcup \mho_{k-1} \\
&\equiv \{ \text{ vertices of all the } \delta^k - \text{cubes that make up } K^{+\delta^k}\} \bigcup \{\mho_{k-1} \bigcap K^c\}
\end{aligned}$$

and

$$\Omega_k = \bigcap_{\theta \in \mho_k} \Omega_\theta.$$

As $\mho_k$ is finite, we have $P(\Omega_k) = 1$ and

$$\text{for any } \omega \in \Omega_k, \quad M_n^k(\omega) = \sup_{\theta \in \mho_k} |\lambda_n(\omega, \theta) - \lambda(\theta)| \to 0. \tag{35}$$

We first establish the connection between $M_n^k(\omega)$ and the upper bound for $\lambda_n(\omega, \theta) - \lambda(\theta)$, over $\theta \in K$, for any given $\omega \in \Omega_k$.

For any fixed $k = 1, 2, \cdots$, each θ in K lies within a δ^k-cube with vertices $\{\theta_i\} \in \mho_k$; it can be written as a convex combination $\sum_i \alpha_i \theta_i$ of those vertices, i.e.

$$\theta = \sum_{\theta_i \in \mho_k} \alpha_i \theta_i, \quad \sum_{\theta_i \in \mho_k} \alpha_i = 1.$$

Then for any given $\omega \in \Omega_k$, convexity of $\lambda_n(\omega, \theta)$ in θ gives

$$\begin{aligned}
\lambda_n(\omega, \theta) &\leq \sum_{\theta_i \in \mho_k} \alpha_i \lambda_n(\omega, \theta_i) \\
&= \sum_{\theta_i \in \mho_k} \alpha_i \{\lambda_n(\omega, \theta_i) - \lambda(\theta_i)\} + \sum_{\theta_i \in \mho_k} \alpha_i \{\lambda(\theta_i) - \lambda(\theta)\} + \lambda(\theta) \\
&\leq M_n^k(\omega) + \max_{\theta_i \in \mho_k} |\lambda(\theta_i) - \lambda(\theta)| + \lambda(\theta).
\end{aligned}$$

Therefore,

$$\lambda_n(\omega, \theta) - \lambda(\theta) \leq M_n^k(\omega) + \epsilon_k. \tag{36}$$

Next we establish the companion lower bound. For any fixed $k = 1, \cdots$, each θ in K lies within a δ^k-cube with a vertex θ_0 in $K \bigcap \mho_k$:

$$\theta = \theta_0 + \sum_{i=1}^{d} \delta_i e_i, \text{ with } |\delta_i| \leq \delta^k, \; i = 1, \cdots, d.$$

Without loss of generality, suppose $\delta_i \geq 0$ for each $i = 1, \cdots, d$. Define

$$\theta_{ik} = \theta_0 - \delta_i' e_i, \quad \text{where } \delta_i' \equiv \min\{c \geq \delta_k : \theta_0 - c e_i \in \mho_k\}, \ i = 1, \cdots, d$$

Note that as $\theta_0 \in K \bigcap \mho_k$, δ_i' must exist and $\delta_i' < 2\delta^k$, for all $i = 1, \cdots, d$.

Write θ_0 as a convex combination of θ and these θ_{ik}:

$$\theta_0 = \frac{\prod_{j=1}^d \delta_j'}{\prod_{j=1}^d \delta_j' + \sum_{j=1}^d \delta_j \prod_{l \neq j} \delta_l'} \theta + \sum_{i=1}^d \frac{\delta_i \prod_{j \neq i} \delta_j'}{\prod_{j=1}^d \delta_j' + \sum_{j=1}^d \delta_j \prod_{l \neq j} \delta_l'} \theta_{ik}.$$

Denote these convex weights by β and $\{\beta_i\}$. As $\delta_j \leq \delta^k \leq \delta_j'$, we have $\beta \geq 1/(d+1)$ and

$$\begin{aligned} \beta \lambda_n(\omega, \theta) &\geq \lambda_n(\omega, \theta_0) - \sum_i \beta_i \lambda_n(\omega, \theta_{ik}) \quad (\text{ convexity of } \lambda_n(\omega, \theta) \text{ in } \theta) \\ &\geq \lambda(\theta_0) - \sum_i \beta_i \lambda(\theta_{ik}) - 2M_n^k(\omega) \quad (\text{ from (35)}) \\ &\geq \lambda(\theta) - \epsilon_k/(d+1) - \sum_i \beta_i[\lambda(\theta) + \epsilon_k/(d+1)] - 2M_n^k(\omega) \\ &= \beta\lambda(\theta) - 2\epsilon_k/(d+1) - 2M_n^k(\omega) \end{aligned}$$

where the third inequality is due to the definition of δ^k and the fact that there exists a cube of side $3\delta^k$ which contains both θ_{ik} and θ_0. As $\beta \geq 1/(d+1)$,

$$\lambda_n(\omega, \theta) - \lambda(\theta) \geq -2\epsilon_k - 2(d+1)M_n^k(\omega).$$

This together with (36) implies that for any $k = 1, 2, \cdots$, there exists some $\Omega_k (\supseteq \Omega_{k+1})$ such that $P(\Omega_k) = 1$ and

$$\forall \omega \in \Omega_k, \ \sup_{\theta \in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \leq (d+1)M_n^k(\omega) + 2k^{-1}.$$

Let $\Omega_0 \equiv \bigcap_{k=1}^\infty \Omega_k$. As Ω_k is a decreasing sequence and $P(\Omega_k) = 1$, we have $P(\Omega_0) = 1$ and for any $\omega \in \Omega_0$,

$$\sup_{\theta \in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \leq (d+1)M_n^k(\omega) + 2k^{-1}, \ \text{for all } k \geq 1. \tag{37}$$

Note that as $n \to \infty$, $M_n^k(\omega) \to 0$ for each fixed k, as in (35). Take limit of both sides of (37)

$$\lim_{n\to\infty} \sup_{\theta \in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \leq \lim_{n\to\infty} M_n^k(\omega) + k^{-1} = k^{-1}, \ \text{for all } k \geq 1.$$

This is equivalent to that with probability 1, $\lim_{n\to\infty} \sup_{\theta \in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \to 0$. ■

We now list a number of facts in the literature that will be used in our proofs later.

Lemma 6.8 [Korolyuk et al, 1989] Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables. With a symmetric kernel $\Phi : X^m \to R$, we consider the U-statistic

$$U_n = \binom{n}{m} \sum_{l \le i_1 < \cdots < i_m \le n} \Phi(X_{i_1}, \cdots, X_{i_m})$$

Let $\theta = E\Phi(X_1, \cdots, X_m) < \infty$ and for $c = 0, 1, \cdots, m$, define

$$\Phi_c(x_1, \cdots, x_c) = E(\Phi(X_1, \cdots, X_m)|X_1 = x_1, \cdots, X_c = x_c), \quad , \Phi_0 = \theta,\ \Phi_m = \Phi$$
$$g_c(x_1, \cdots, x_c) = \sum_{d=0}^{c} (-1)^{c-d} \sum \sum_{l \le j_1 < \cdots < j_d \le c} \Phi_d(x_{j1}, \cdots, x_{jd}), \quad \sigma_1^2 = Eg_1^2(X_1)$$

Suppose $\sigma_1^2 > 0$ and for all $c = 1, \cdots, m$, $Eg_c^{2c/(2c-1)} < \infty$. The with probability 1,

$$\limsup_{n\to\infty} \frac{n^{1/2}(U_n - \theta)}{(2m^2\sigma_1^2 \log\log n)^{1/2}} = 1 \quad \blacksquare$$

Lemma 6.9 [Berbee's Lemma] Let (X, Y) be a $R^d \times R^{d'}-$valued random vector. Then there exists a $R^{d'}-$valued random vector Y^* which has the same distribution as Y and

$$Y^* \text{ is independent of } X;\ P(Y^* \ne Y) = \beta(\sigma(X), \sigma(Y)) \tag{38}$$

where $\sigma(X)$ and $\sigma(Y))$ are the $\sigma-$algebra generated by X and Y respectively, and

$$\beta[\sigma(X), \sigma(Y)] = E \sup_{A \in \sigma(Y)} |P(A) - P(A|\sigma(X))|$$

Lemma 6.10 $\beta[\sigma(X_1, Y_1), \sigma(\hat{a}_j, \hat{b}_j)] = O\{(nh/\log^3 n)^{-1/4}\}$

Proof By the definition,

$$\beta[\sigma(X_1, Y_1), \sigma(\hat{a}_j, \hat{b}_j)] = E \sup_{A \in \sigma(\hat{a}_j, \hat{b}_j)} |P(A) - P(A|\sigma(X_1, Y_1))|$$

According to results in Welsh (1996), $[(\hat{a}_j - E\hat{a}_j)/\sigma_1, (\hat{b}_j - E\hat{b}_j)/\sigma_2]$ are asymptotically normal, where $\sigma_1 \equiv \{\text{Var}\hat{a}_j\}^{1/2} = O\{(nh)^{-1/2}\}$ and $\sigma_2 \equiv \{\text{Var}\hat{b}_j\}^{1/2} = O\{(nh^3)^{-1/2}\}$. Let $\tau_n = (nh/\log n)^{-3/4}$ and rewrite (16) as

$$\hat{a}_j = E\hat{a}_j + \frac{1}{nh} \sum_{i=2}^{n} K_{ij}^{\vartheta} \varphi_{ij} + \frac{1}{nh} K_{1j} \varphi_1(X_1, Y_1) + O(\tau_n),$$

$$\hat{b}_j = E\hat{b}_j + \frac{1}{nh^2} \sum_{i=2}^{n} \tilde{\varphi}_{ij} + \frac{1}{nh^2} \tilde{\varphi}_{1j} + O\{\tau_n/h\}. \tag{39}$$

Note that $\varphi_{ij},\ \tilde{\varphi}_{ij},\ i = 1, \cdots, n$ are two sequences of zero-mean i.i.d. bounded random variables defined in (18), whence

$$
\begin{aligned}
P\{\hat{a}_j \le t_1, \hat{b}_j \le t_2 | Y_1, X_1\} &\le P[\hat{a}_j \le C\tau_n + t_1, \hat{b}_j \le C\tau_n/h + t_2] \\
&\le P\Big[(\hat{a}_j - E\hat{a}_j)/\sigma_1 \le (t_1 - E\hat{a}_j + C\tau_n)/\sigma_1, \\
&\quad (\hat{b}_j - E\hat{b}_j)/\sigma_2 \le (t_2 - E\hat{b}_j + C\tau_n/h)/\sigma_2\Big] \\
&= P[\hat{a}_j \le t_1, \hat{b}_j \le t_2] + C(nh)^{1/2}\tau_n, \\
P\{\hat{a}_j \ge t_1, \hat{b}_j \ge t_2 | Y_1, X_1\} &\ge P[\hat{a}_j \ge t_1 - C\tau_n, \hat{b}_j \ge t_2 - C\tau_n/h] \\
&\ge P\Big[(\hat{a}_j - E\hat{a}_j)/\sigma_1 \ge (t_1 - E\hat{a}_j - C\tau_n)/\sigma_1, \\
&\quad (\hat{b}_j - E\hat{b}_j)/\sigma_2 \ge (t_2 - E\hat{b}_j - C\tau_n/h)/\sigma_2\Big] \\
&= P[\hat{a}_j \ge t_1, \hat{b}_j \ge t_2] - C(nh)^{1/2}\tau_n.
\end{aligned}
$$

Therefore,

$$
|P\{\hat{a}_j \le t_1, \hat{b}_j \le t_2 | Y_1, X_1\} - P\{\hat{a}_j \le t_1, \hat{b}_j \le t_2\}| \le C(nh)^{-1/2}\tau_n = O\{(nh/\log^3 n)^{-1/4}\}.
$$

Lemma 6.11 *Under the assumptions (A1)–(A5), we have*

$$
E\Phi_n(\theta) = \delta_\theta^\top ER_{n1}(\theta) + \delta_\theta^\top G_{n\vartheta}\delta_\theta + o(n^2h|\delta_\theta|^2).
$$

Proof Apparently it suffices to show that

$$
\begin{aligned}
& EK_{ij}^\vartheta\{\rho(Y_1 - \hat{a}_j - \hat{b}_j\theta^\top X_{1j}) - \rho(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})\} \\
= \;& \delta_\theta^\top E[K_{ij}^\vartheta\varphi(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})\hat{b}_j X_{1j}] + \delta_\theta^\top E[K_{ij}^\vartheta X_{1j}X_{1j}^\top g(X_1)\hat{b}_j^2]\delta_\theta + o(|\delta_\theta|^2).
\end{aligned}
$$

By the continuity of $E[\rho(Y_1 - \hat{a}_j - t\hat{b}_j)|\mathcal{X}]$ in t, where $\mathcal{X} = \sigma(X_1, \cdots, X_n)$, we have

$$
\begin{aligned}
& E\{\rho(Y_1 - \hat{a}_j - \hat{b}_j\theta^\top X_{1j}) - \rho(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})|\mathcal{X}\} \\
= \;& \delta_\theta^\top X_{1j}E[\varphi(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})\hat{b}_j|\mathcal{X}] + \delta_\theta^\top X_{1j}X_{1j}^\top\delta_\theta\partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t|_{t=X_{1j}^\top\theta_0} \\
& +\delta_\theta^\top X_{1j}X_{1j}^\top\delta_\theta\Big[\partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t|_{t=X_{1j}^\top\theta_0} - \partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t|_{t=t^*}\Big]
\end{aligned}
$$

where t^* is some value between $\theta^\top X_{1j}$ and $\theta_0^\top X_{1j}$. Taking expectations of both sides, we have

$$
\begin{aligned}
& EK_{ij}^\vartheta\{\rho(Y_1 - \hat{a}_j - \hat{b}_j\theta^\top X_{1j}) - \rho(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})\} \qquad (40) \\
= \;& \delta_\theta^\top E[K_{ij}^\vartheta\varphi(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^\top X_{1j})\hat{b}_j X_{1j}] + \delta_\theta^\top(\Delta_1 + \Delta_2)\delta_\theta \\
\Delta_1 = \;& E\{K_{ij}^\vartheta X_{1j}X_{1j}^\top\partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t|_{t=X_{1j}^\top\theta_0}\} \\
\Delta_2 = \;& E\{K_{ij}^\vartheta X_{1j}X_{1j}^\top\partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t|_{t=X_{1j}^\top\theta_0}\} - \Delta_1
\end{aligned}
$$

where t^* is some value between $\theta^\top X_{1j}$ and $\theta_0^\top X_{1j}$.

To study Δ_1, we need to compute $\partial[E\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}\}]/\partial t$. To this end, we apply Lemma 6.9 and Lemma 6.10. Suppose $[\tilde{a}_j, \tilde{b}_j]$ has the same distribution as $[\hat{a}_j, \hat{b}_j]$, but is independent of (Y_1, X_1) and $P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]) = O\{(nh/\log^3 n)^{-1/4}\}$. Thus for any $\delta \to 0$,

$$
\begin{aligned}
& E[\varphi(Y_1 - \hat{a}_j - \hat{b}_j(t+\delta))\hat{b}_j|\mathcal{X}] - E[\varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}] \\
= \; & E[\varphi\{Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)\}\tilde{b}_j] - E[\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j t)\tilde{b}_j|\mathcal{X}] \\
& + E[\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j(t+\delta)) - \varphi(Y_1 - \hat{a}_j - \hat{b}_j t)\}\hat{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}] \\
& - E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)) - \varphi(Y_1 - \tilde{a}_j - \tilde{b}_j t)\}\tilde{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}] \\
\equiv \; & \mathrm{T}_1 + \mathrm{T}_2 + \mathrm{T}_3
\end{aligned} \tag{41}
$$

Based on the definition of $G_1(s; X)$, since Y_1 is independent of $[\tilde{a}_j, \tilde{b}_j]$, we have

$$
\begin{aligned}
\mathrm{T}_1 &= E[\{G_1(a_1 - \tilde{a}_j - \tilde{b}_j(t+\delta); X_1) - G_1(a_1 - \tilde{a}_j - \tilde{b}_j t; X_1)\}\tilde{b}_j|\mathcal{X}] \\
&= \delta E[G_2(a_1 - \tilde{a}_j - \tilde{b}_j t; X_1)\tilde{b}_j^2|\mathcal{X}] + o(\delta),
\end{aligned} \tag{42}
$$

where the last equality follows from the continuity of $G_1(t; X)$ in t.

Next, we show that $\mathrm{T}_2 = o(\delta)$. As we mentioned in the proof of Lemma 6.10, $[v_1, v_2] \equiv [(\hat{a}_j - E\hat{a}_j)/\sigma_1, (\hat{b}_j - E\hat{b}_j)/\sigma_2]$ are asymptotically normal, where

$$
\sigma_1 \equiv \{\mathrm{Var}\hat{a}_j\}^{1/2} = O\{(nh)^{-1/2}\}, \; \sigma_2 \equiv \{\mathrm{Var}\hat{b}_j\}^{1/2} = O\{(nh^3)^{-1/2}\}.
$$

Similarly construct $[\tilde{v}_1, \tilde{v}_2]$ from $\tilde{a}_j$ and $\tilde{b}_j$. Without loss of generality, consider a small $\delta (> 0)$. It is easy to understand that the conditional probability density function of Y_1 given $[v_1, v_2]$ is uniformly bounded. Therefore, for any given values of $\hat{a}_j$ and $\hat{b}_j$ (equivalently v_1 and v_2),

$$
|E\{\varphi(Y_i - \hat{a}_j - \hat{b}_j(t+\delta)) - \varphi(Y_i - \hat{a}_j - \hat{b}_j t)|v_1, v_2\}| \le C\delta|\hat{b}_j|.
$$

Let $f(\tilde{v}_1, \tilde{v}_2|v_1, v_2)$ be the conditional probability density function of $(\tilde{v}_1, \tilde{v}_2)$ given (v_1, v_2), and

$$
g(v_1, v_2) = \int_{[\tilde{v}_1, \tilde{v}_2] \neq [v_1, v_2]} f(\tilde{v}_1, \tilde{v}_2|v_1, v_2) d\tilde{v}_1 d\tilde{v}_2.
$$

As $\int f(v_1, v_2) g(v_1, v_2) dv_1 dv_2 = P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]) = O\{(nh/\log^3 n)^{-1/4}\}$, we have

$$
|\mathrm{T}_2| \le C\delta \int \hat{|}b_j| f(v_1, v_2) g(v_1, v_2) dt_1 dt_2 = o(\delta).
$$

Similarly we can show that $\mathrm{T}_3 = o(\delta)$. This together with (41) and (42) yields

$$\partial[E\varphi(Y_i - \hat{a}_j - \hat{b}_j t)\hat{b}_j|\mathcal{X}]/\partial t = E[G_2(a_1 - \tilde{a}_j - \tilde{b}_j t; X_1)\tilde{b}_j^2|\mathcal{X}]. \tag{43}$$

Apply this result to Δ_1 and Δ_2, we have

$$\Delta_1 = E[K_{ij}^{\vartheta} X_{1j} X_{1j}^{\top} G_2(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0; X_1)\tilde{b}_j^2], \quad \Delta_2 = O(\delta_\theta).$$

Plugging this into (40) leads to

$$\begin{aligned}
& EK_{ij}^{\vartheta}\{\rho(Y_1 - \hat{a}_j - \hat{b}_j\theta^{\top}X_{1j}) - \rho(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^{\top}X_{1j})\} \\
= \; & \delta_\theta^{\top} E[K_{ij}^{\vartheta}\varphi(Y_1 - \hat{a}_j - \hat{b}_j X_{1j}^{\top}\theta_0)\hat{b}_j X_{1j}] + \delta_\theta^{\top} E[K_{ij}^{\vartheta} X_{1j} X_{1j}^{\top} G_2(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0; X_1)\tilde{b}_j^2]\delta_\theta + o(|\delta_\theta|^2) \\
= \; & \delta_\theta^{\top} E[K_{ij}^{\vartheta}\varphi(Y_1 - \hat{a}_j - \hat{b}_j\theta_0^{\top}X_{1j})\hat{b}_j X_{1j}] + \delta_\theta^{\top} E[K_{ij}^{\vartheta} X_{1j} X_{1j}^{\top} g(X_1) b_j^2]\delta_\theta + o(|\delta_\theta|^2)
\end{aligned}$$

where the last equality follows from the continuity of $G_2(t; X_1)$ in t and (19). ■

Lemma 6.12 *Define* $Z_{ij} = K_{ij}^{\vartheta}\hat{b}_j X_{ij}\{\varphi(Y_{ij}) - \varphi(\varepsilon_i)\}$. *Then*

$$h^{-1}E_i Z_{ij} = -\delta_\vartheta^{\top} b_j^2\{(\nu/\mu)_\vartheta(X_j) - X_j\}\{\nu_\vartheta(X_j) - X_j\mu_\vartheta(X_j)\}^{\top} + o(|\delta_\vartheta| + n^{-1/2}), \tag{44}$$

$$\sum_{i,j}(Z_{ij} - E_i Z_{ij}) = o(n^2 h\delta_\vartheta), \tag{45}$$

$$(nh)^{-1}\sum_i K_{ij}^{\vartheta}\varphi(\varepsilon_i)(\hat{b}_j - b_j)X_{ij} = o(n^{-1/2}) + O\{\delta_\vartheta(nh/\log n)^{-1/2}\} \tag{46}$$

uniformly in ϑ.

Proof Once again we apply Lemma 6.9 and suppose $[\tilde{a}_j, \tilde{b}_j]$ has the same distribution as $[\hat{a}_j, \hat{b}_j]$ and is independent of (X_1, Y_1). By Lemma 6.10, $P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}) = O\{(nh/\log^3 n)^{-1/4}\}$. Recall $\mathcal{X} = \sigma(X_1, \cdots, X_n)$. Note that $E_1 Z_{1j} = E[K_{1j}X_{1j}(\mathrm{T}_1 - \mathrm{T}_2 + \mathrm{T}_3)]$, where

$$\begin{aligned}
& E[\{\varphi(Y_1 - \hat{a}_j - X_{1j}^{\top}\theta_0\hat{b}_j) - \varphi(\varepsilon_1)\}\hat{b}_j|\mathcal{X}] = \mathrm{T}_1 - \mathrm{T}_2 + \mathrm{T}_3, \\
& \mathrm{T}_1 = E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0) - \varphi(\varepsilon_1)\}\tilde{b}_j|\mathcal{X}] \\
& \mathrm{T}_2 = E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0) - \varphi(\varepsilon_1)\}\tilde{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}] \\
& \mathrm{T}_3 = E[\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j X_{1j}^{\top}\theta_0) - \varphi(\varepsilon_1)\}\hat{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}].
\end{aligned}$$

Similar to (42), we can conclude that

$$\begin{aligned}
\mathrm{T}_1 \; &= \; E[\{G_1(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0; X_1) - G_1(0; X_1)\}\tilde{b}_j|\mathcal{X}] \qquad (47) \\
&= \; g(X_1)E\{\tilde{b}_j(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0)|\mathcal{X}\} + O[E\{(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^{\top}\theta_0)^2|\mathcal{X}\}].
\end{aligned}$$

Using the results on the asymptotic bias and variance of $(\tilde a_j, \tilde b_j)$ in (19), we can see that

$$E\{K_{1j}^{\vartheta}(a_1 - \tilde a_j - \tilde b_j X_{1j}^{\top}\theta_0)^2\} = O(h\delta_{\vartheta}^2 + n^{-1}),$$

Next we deal with the first term in (47). Using (16),

$$\begin{aligned}
& a_1 - \tilde a_j - \tilde b_j X_{1j}^{\top}\theta_0 = a_1 - a_j + a_j - \tilde a_j - \tilde b_j X_{1j}^{\top}\theta_0 \\
& = \frac{1}{2}m''(X_j^{\top}\theta_0)\{(X_{1j}^{\top}\theta_0)^2\} - \frac{1}{2}m''(X_j^{\top}\theta_0)h^2 + O\{(X_{1j}^{\top}\theta_0)^3\} \\
& -b_j\delta_{\vartheta}^{\top}\{(\nu/\mu)_{\vartheta}(X_j) - X_j\} - b_j\delta_{\vartheta}^{\top}\{(\mu\nu' - \mu'\nu)/\mu^2\}_{\vartheta}(X_j)X_{1j}^{\top}\theta_0 \\
& -h^2[\frac{1}{2}m''(X_j^{\top}\theta_0)\{(f\mu)'/(fg)\}_{\vartheta}(X_j) + \frac{1}{6}m^{(3)}(X_j^{\top}\theta_0)(f\mu)_{\vartheta}(X_j)]X_{1j}^{\top}\theta_0 \\
& +\{gf\}_{\vartheta}^{-1}(X_j)\frac{1}{nh}\sum_{i=1}^{n}\varphi_{ij} - \{gf\}_{\vartheta}^{-1}(X_j)\{\frac{1}{nh^2}\sum_{i=1}^{n}\tilde\varphi_{ij}\}X_{1j}^{\top}\theta_0 \\
& +O\{(nh/\log n)^{-3/4}(1+\delta_{\vartheta}/h) + h^3\} \qquad (48)
\end{aligned}$$

where $\varphi_{ij},\ \tilde\varphi_{ij}$ are zero-mean IID random variables

$$\begin{aligned}
E[K_{1j}^{\vartheta}X_{1j}\mathrm{T}_1] &= E[K_{1j}^{\vartheta}g(X_1)X_{1j}\tilde b_1(a_1 - \tilde a_j - \tilde b_j X_{1j}^{\top}\theta_0)] + o(h|\delta_{\vartheta}| + n^{-1/2}h) \qquad (49)\\
&= -h\delta_{\vartheta}^{\top}b_j^2\{(\nu/\mu)_{\vartheta}(X_j) - X_j\}\{\nu_{\vartheta}(X_j) - X_j\mu_{\vartheta}(X_j)\} + o(h|\delta_{\vartheta}| + hn^{-1/2})
\end{aligned}$$

uniformly in ϑ, where (19) is used in the last step.

As $P([\tilde a_j, \tilde b_j] \neq [\hat a_j, \hat b_j]) = O\{(nh/\log^3 n)^{-1/4}\}$, we have similar to T_2 in (41),

$$E[K_{1j}^{\vartheta}X_{1j}\mathrm{T}_2] = o(n^{-1/2}h) + o(h\delta_{\vartheta}), \quad E[K_{1j}^{\vartheta}X_{1j}\mathrm{T}_2] = o(n^{-1/2}h) + o(h\delta_{\vartheta})$$

uniformly in ϑ. This together with (49) yields (44).

To prove (45), first note that

$$\begin{aligned}
\varphi(Y_i - \hat a_j - \hat b_j\theta_0^{\top}X_{ij}) - \varphi(\varepsilon_i) &= [\varphi(Y_i - \hat a_j - \hat b_j\theta_0^{\top}X_{ij}) - \varphi(Y_i - a_j - b_j\theta_0^{\top}X_{ij})] \\
&\quad + [\varphi(Y_i - a_j - b_j\theta_0^{\top}X_{ij}) - \varphi(\varepsilon_i)].
\end{aligned}$$

Let $\tilde Z_{ij} = K_{ij}^{\vartheta}X_{ij}\{\varphi(Y_i - a_j - b_j\theta_0^{\top}X_{ij}) - \varphi(\varepsilon_i)\}$. By Lemma 6.14, it suffices to show that

$$\sum_{i,j} b_j(\tilde Z_{ij} - E\tilde Z_{ij}) = o(n^2h\delta_{\vartheta}) \qquad (50)$$

$$\sum_{j}(\hat b_j - b_j)\sum_{i}\tilde Z_{ij} = o(n^2h\delta_{\vartheta}). \qquad (51)$$

Due to Borel-Cantelli Lemma, (50) can be further reduced to, for any $\epsilon > 0$

$$nP\{|\sum_{i} b_j(\tilde Z_{ij} - E\tilde Z_{ij})| \geq \epsilon nh\delta_{\vartheta}\} \text{ is summable over } n, \qquad (52)$$

which follows from the facts that $\tilde{Z}_{ij}$ is bounded, $E\tilde{Z}_{ij}^2 = O(h^3 + h\delta_\vartheta^2)$ and Bernstein's inequality,

$$P\{|\sum_i (\tilde{Z}_{ij} - E\tilde{Z}_{ij})| \geq \epsilon nh\delta_\vartheta\} \leq C\exp\Big\{-\frac{\epsilon^2 n^2 h^2 \delta_\vartheta^2}{nh^3 + nh\delta_\vartheta^2 + \epsilon nh\delta_\vartheta}\Big\} = o(n^{-2}).$$

To prove (51), we again use the expansion of $\hat{b}_j - b_j$ given in (16), i.e.

$$\begin{aligned} \hat{b}_j - b_j &= h^2\Big[\frac{1}{2}m''(X_j^\top\theta_0)\{(f\mu)'/(fg)\}_\vartheta(X_j) + \frac{1}{6}m^{(3)}(X_j^\top\theta_0)\{(f\mu)/(fg)\}_\vartheta(X_j)\Big] \\ &\quad + b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + \frac{1}{nh^2}\sum_{i=1}^n \tilde{\varphi}_{ij} + O\{(nh/\log n)^{-3/4}/h\} \end{aligned}$$

where $E\tilde{\varphi}_{ij} = 0$. If we denote by $C(X_j)$ the determinstic(bias) term in $\hat{b}_j - b_j$, it is easy to see that $\sum_{i,j} C(X_j)\tilde{Z}_{ij} = o(n^2h\delta_\vartheta)$. For the stochastic part, write

$$\sum_{j,i,l}\tilde{Z}_{ij}\tilde{\varphi}_{lj} = \sum_{i,j}\tilde{Z}_{ij}\tilde{\varphi}_{ij} + \sum_{j,i\neq l}\tilde{Z}_{ij}\tilde{\varphi}_{lj} \tag{53}$$

We focus on the second term, as the first term is relatively negligible. Let $c \equiv E\tilde{Z}_{ij} = O(h^3 + h\delta_\vartheta^2)$, whence the second term in (53) is $(nh^2)^{-1}\sum_j(\mathrm{T}_{1j} + c\mathrm{T}_{2j})$, where

$$\mathrm{T}_{1j} = \sum_{i<l}\{\tilde{\varphi}_{lj}(\tilde{Z}_{ij} - c) + \tilde{\varphi}_{ij}(\tilde{Z}_{lj} - c)\}, \quad \mathrm{T}_{2j} = \sum_{i<l}(\tilde{\varphi}_{lj} + \tilde{\varphi}_{ij}).$$

By the second statement in Lemma 6.1 in Xia(2007), replacing θ there with $(\vartheta^\top, X_j^\top)^\top$, we know that with probabiltity 1, $\mathrm{T}_{1j} = O\{n\log n(h^3 + h\delta_\vartheta^2)^{1/2}\}$ uniformly in ϑ and j. On the other hand, by law of the iterated logarithm for U-statistics in Korolyuk et al (Lemma 6.8), $\sum_j \mathrm{T}_{2j} = n^{3/2}(h\log\log n)^{1/2}$ *a.s.* Since $c = O(h^3 + h\delta_\vartheta^2)$, we have

$$\frac{1}{nh^2}\sum_j(\mathrm{T}_{1j} + c\mathrm{T}_{2j}) = \frac{1}{nh^2}O\{n^2\log n(h^3 + h\delta_\vartheta^2)^{1/2} + n^{3/2}\log n)(h^3 + h\delta_\vartheta^2)\} = o(n^2h\delta_\vartheta)$$

Proof of (46) can be done in exactly the same manner as (51). ■

The proof of (26) consists of the following two Lemmas.

Lemma 6.13 *Let* $R_{n2}^*(\theta) = \sum_{i,j} K_{ij}^\vartheta\Big[\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_{ij}) - \delta_\theta^\top\varphi(Y_i - a_j - b_jX_{ij}^\top\theta_0)\hat{b}_jX_{ij}\Big]$. *Then with probability* 1, *we have*

$$(n^2ha_{n\vartheta}^2)^{-1}[R_{n2}^*(\theta) - ER_{n2}^*(\theta)] = o(1). \tag{54}$$

uniformly in ϑ.

Proof Define $X_{ix} = X_i - x$, $\mu_{ix} = (1, X_{ix}^\top)^\top$, $K_{ix} = K(X_{ix}^\top \vartheta / h)$, $\beta(x) = [m(\theta_0^\top x), m'(\theta_0^\top x)\theta_0^\top]^\top$ and $\varphi_{ni}(x; t) = \varphi(Y_i; \mu_{ix}^\top \beta(x) + t)$. For any α, $\beta \in \mathcal{R}^{d+1}$, let

$$\begin{aligned}
\Phi_{ni}(x; \alpha, \beta) &= K_{ix}\Big[\rho\{Y_i; \mu_{ix}^\top(\alpha + \beta + \beta(x))\} - \rho\{Y_i; \mu_{ix}^\top(\beta + \beta(x))\}) - \varphi_{ni}(x; 0)\mu_{ix}^\top \alpha\Big] \\
&= K_{ix} \int_{\mu_{ix}^\top \beta}^{\mu_{ix}^\top(\alpha+\beta)} \{\varphi_{ni}(x; t) - \varphi_{ni}(x; 0)\} dt
\end{aligned}$$

and $R_{ni}(x; \alpha, \beta) = \Phi_{ni}(x; \alpha, \beta) - E\Phi_{ni}(x; \alpha, \beta)$. Apparently,

$$K_{ij}^\vartheta \Big[\rho(Y_i - \hat{a}_j - \hat{b}_j \theta^\top X_{ij}) - \rho(Y_{ij}) - \delta_\theta^\top \varphi(Y_i - a_j - b_j X_{ij}^\top \theta_0)\hat{b}_j X_{ij}\Big] \equiv \Phi_{ni}(X_j; \alpha, \beta)$$

with $\alpha = [0, \hat{b}_j \delta_\theta^\top]^\top$ and $\beta = [\hat{a}_j - a_j, (\hat{b}_j - b_j)\theta_0^\top]^\top$. Let $[a_x, b_x] \equiv [m(\theta_0^\top x), m'(\theta_0^\top x)]$ and $\mathcal{D}$ be any compact subset of the support of X. For any $M > 0$ and $\vartheta \in \Theta_n$, define

$$\begin{gathered}
M_{n1}^\vartheta = C a_{n\vartheta}, \quad M_{n2}^\vartheta = C\{|\delta_\vartheta| + (nh/\log n)^{-1/2}\}, \\
M_{n3}^\vartheta = C\{|\delta_\vartheta| + (nh/\log n)^{-1/2}/h\}, \quad B_n^{(1)} = \{\alpha \in R^{d+1} | \alpha = [0, \alpha_1^\top]^\top, |\alpha_1| \le M_{n1}^\vartheta\}, \\
B_n^{(2)} = \{\beta \in R^{d+1} | \beta = [b_1, b_2\theta_0^\top]^\top, |b_1| \le M_{n2}^\vartheta, |b_2| \le M_{n3}^\vartheta\}.
\end{gathered}$$

As $|\hat{b}_j \delta_\theta| \le C a_{n\vartheta}$, $|\hat{a}_j - a_j| = O\{|\delta_\vartheta| + (nh/\log n)^{-1/2}\}$ and $|(\hat{b}_j - b_j)| = O\{|\delta_\vartheta| + (nh/\log n)^{-1/2}/h\}$, (54) will follow if for any $\epsilon > 0$

$$\sup_{x \in \mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(x; \alpha, \beta)| \le \epsilon d_n \ a.s., \quad d_n = nha_{n\vartheta}^2 \tag{55}$$

This is done in a similar style as Lemma 4.2 in Kong et al(2008). Cover $\mathcal{D}$ by a finite number T_n of cubes $\mathcal{D}_k = \mathcal{D}_{n,k}$ with side length $l_n = O\{h(nh/\log n)^{-1/4}\}$ and centers $x_k = x_{n,k}$. Write

$$\begin{aligned}
\sup_{x \in \mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(x; \alpha, \beta)| \le & \max_{1 \le k \le \mathrm{T}_n} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n R_{ni}(x_k; \alpha, \beta)\Big| \\
& + \max_{1 \le k \le \mathrm{T}_n} \sup_{x \in \mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{\Phi_{ni}(x_k; \alpha, \beta) - \Phi_{ni}(x; \alpha, \beta)\Big\}\Big| \\
& + \max_{1 \le k \le \mathrm{T}_n} \sup_{x \in \mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{E\Phi_{ni}(x_k; \alpha, \beta) - E\Phi_{ni}(x; \alpha, \beta)\Big\}\Big| \\
\equiv & Q_1 + Q_2 + Q_3.
\end{aligned}$$

In Lemma 6.15, we will prove that $Q_2 = o(d_n),\ a.s.$, whence $Q_3 \le EQ_2 = o(d_n)$. It remains to show that $Q_1 \le \epsilon d_n/3\ a.s.$, which can be done following a similar proof style as in Lemma 4.2 in Kong et al (2008).

Partition $B_n^{(i)},\ i = 1, 2$ into a sequence of sub rectangles $D_1^{(i)}, \cdots, D_{J_1}^{(i)},\ i = 1, 2$, such that for all $1 \le j_1 \le J_1 \le M^{d+1}$ $(M = \epsilon^{-1})$ and for all$\alpha, \alpha' \in D_{j_1}^{(1)}$, we have $|\alpha - \alpha'| \le M_{n1}^{\vartheta}/M$; for all $\beta = [b_1, b_2\theta_0^\top]^\top, \beta' = [b_1', b_2'\theta_0^\top]^\top \in D_{j_1}^{(2)}$, we have $|b_1 - b_1'| \le M_{n2}^{\vartheta}/M, |b_2 - b_2'| \le M_{n3}^{\vartheta}/M$. Choose a point $\alpha_{j_1} \in D_{j_1}^{(1)}$ and $b_{k_1} \in D_{k_1}^{(2)}$, $1 \le j_1, k_1 \le J_1$. Then for any x,

$$\begin{aligned}\sup_{\substack{\alpha \in B_n^{(1)} \\ \beta \in B_n^{(2)}}} |\sum_i R_{ni}(x; \alpha, \beta)| \quad \le \quad & \max_{1 \le j_1, k_1 \le J_1} \sup_{\substack{\alpha \in D_{j_1}^{(1)}, \\ \beta \in D_{k_1}^{(2)}}} |\sum_{i=1}^n \{R_{ni}(x; \alpha_{j_1}, b_{k_1}) - R_{ni}(x; \alpha, \beta)\}| \\ & + \max_{1 \le j_1, k_1 \le J_1} |\sum_{i=1}^n R_{ni}(x; \alpha_{j_1}, \beta_{k_1})| = H_{n1} + H_{n2}. \end{aligned} \tag{56}$$

We first show that any $\epsilon > 0$

$$\mathrm{T}_n P\Big\{H_{n2} \ge \frac{\epsilon d_n}{2}\Big\} \le \mathrm{T}_n J_1^2 P\Big\{|\sum_{i=1}^n R_{ni}(x; \alpha_{j_1}, \beta_{k_1})| \ge \frac{\epsilon d_n}{3}\Big\} = O(n^{-a}), \tag{57}$$

for some $a > 1$. By Bernstein's Inequality and the fact that $|R_{ni}(x; \alpha_{j_1}, \beta_{k_1})| \le C a_{n\vartheta}$ and $\mathrm{Var}\{R_{ni}(x; \alpha_{j_1}, \beta_{k_1})\} = O[nha_{n\vartheta}^2\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}]$, we have

$$T_n J_1^2 P\Big\{|\sum_{i=1}^n R_{ni}(x; \alpha_{j_1}, \beta_{k_1})| \ge \frac{\epsilon d_n}{3}\Big\} = T_n J_1^2 \exp[-\epsilon^2 nha_{n\vartheta}\{1 + a_{n\vartheta}(nh/\log n)^{1/2})\}] = O(n^{-a}),$$

for some $a > 1$. Therefore, (57) holds.

We next consider H_{n1}. For each $j_1 = 1, \cdots, J_1$ and $i = 1, 2$, partition each rectangle $D_{j_1}^{(i)}$ further into a sequence of subrectangles $D_{j_1,1}^{(i)}, \cdots, D_{j_1,J_2}^{(i)}$. Repeat this process recursively as follows. Suppose after the lth round, we get a sequence of rectangles $D_{j_1,j_2,\cdots,j_l}^{(i)}$ with $1 \le j_k \le J_k,\ 1 \le k \le l$, then in the $(l+1)$th round, each rectangle $D_{j_1,j_2,\cdots,j_l}^{(i)}$ is partitioned into a sequence of subrectangles $\{D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}, 1 \le j_l \le J_l\}$ such that for all $1 \le j_{l+1} \le J_{l+1}$ and for all $a, a' \in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}$, we have $|a - a'| \le M_{n1}^{\vartheta}/M^{l+1}$; and for all $\beta = [b_1, b_2\theta_0^\top]^\top, \beta' = [b_1', b_2'\theta_0^\top]^\top \in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(2)}, |b_1 - b_1'| \le M_{n2}^{\vartheta}/M^{l+1}, |b_2 - b_2'| \le M_{n3}^{\vartheta}/M^{l+1}$, where $J_{l+1} \le M^{d+1}$. Repeat this process after the $(\mathrm{L}_n + 2)$th round, with L_n being the largest integer such that

$$n(2/M)^{L_n} > d_n/M_{n2}^{\vartheta}. \tag{58}$$

Let $D_l^{(i)},\ i = 1, 2$, denote the set of all subrectangles of $D_0^{(i)}$ after the lth round of partition and a typical element $D_{j_1,j_2,\cdots,j_l}^{(i)}$ of $D_l^{(i)}$ is denoted as $D_{(j_l)}^{(i)}$. Choose a point $\alpha_{(j_l)} \in D_{(j_l)}^{(1)}$ and

$\beta_{(j_l)} \in D^{(2)}_{(j_l)}$. Define

$$V_l = \sum_{\substack{(j_{l+1})\\(k_{l+1})}} P\Big\{\Big|\sum_{i=1}^n \{R_{ni}(x;\alpha_{(j_l)},\beta_{(k_l)}) - R_{ni}(x;\alpha_{(j_{l+1})},\beta_{(k_{l+1})})\}\Big| \geq \frac{\varepsilon d_n}{2^{l+1}}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 1,$$

$$Q_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\sup_{\substack{\alpha\,\in\, D^{(1)}_{(j_l)},\\ \beta\,\in\, D^{(2)}_{(k_l)}}} \Big|\sum_{i=1}^n \{R_{ni}(x;\alpha_{(j_l)},\beta_{(k_l)}) - R_{ni}(x;\alpha,\beta)\}\Big| \geq \frac{\varepsilon d_n}{2^l}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 2.$$

Then $Q_l \leq V_l + Q_{l+1},\ 1 \leq l \leq L_n + 1$. On the other hand, it is easy to see that for any $\alpha \in D^{(1)}_{(j_{L_n+2})}$ and $\beta \in D^{(2)}_{(k_{L_n+2})}$,

$$n|R_{ni}(x;\alpha_{(j_{L_n+2})},\beta_{(k_{L_n+2})}) - R_{ni}(x;\alpha,\beta)| \leq nM^{\vartheta}_{n2}/M^{L_n+2} \leq \epsilon d_n/2^{L_n+2}$$

due to the choice of L_n specified in (58). Therefore, $Q_{L_n+2} = 0$ and it remains to show that

$$T_n P\{H_{n1} \geq \frac{\epsilon d_n}{2}\} \leq T_n J_1^2 Q_1 \leq T_n J_1^2 \sum_{l=1}^{L_n+1} V_l = O(n^{-a}),\ \text{for some } a > 1. \tag{59}$$

To find upper bound for $V_l,\ 1 \leq l \leq L_n + 1$, we again apply Bernstein's inequality. As

$$\begin{aligned} & |R_{ni}(x;\alpha_{(j_l)},\beta_{(k_l)}) - R_{ni}(x;\alpha_{(j_{l+1})},\beta_{(k_{l+1})})| \\ \leq\ & C\{|\alpha_{(j_l)} - \alpha_{(j_{l+1})}| + |\beta_{(k_l)} - \beta_{(k_{l+1})}|(\delta_\vartheta + h)\} \equiv M^{\vartheta}_{n2}/M^l, \end{aligned}$$

$$E|R_{ni}(x;\alpha_{(j_l)},\beta_{(k_l)}) - R_{ni}(x;\alpha_{(j_{l+1})},\beta_{(k_{l+1})})|^2 \leq h(M^{\vartheta}_{n2})^3/M^l,$$

we have

$$V_l \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big)\exp[-\varepsilon^2 nh\{1 + a_{n\vartheta}(nh/\log n)^{1/2}\}],$$

and (59) thus holds. This together with (57) completes the proof. ■

Lemma 6.14 *Let* $Z_{ij} = K_{ij}[\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})]\hat{b}_j X_{ij}$. *Then*

$$\sum_{i,j} Z_{ij} - EZ_{ij} = o(n^2 h a_{n\vartheta}). \tag{60}$$

Proof As $\hat{a}_j - a_j = O(a_{n\vartheta})$, $(\hat{b}_j - b_j) = O\{a_{n\vartheta} + (nh/\log n)^{1/2}/h\}$ and for any $\epsilon > 0$,

$$P\Big\{|\sum_{i,j} Z_{ij} - EZ_{ij}| \geq \epsilon n^2 h a_{n\vartheta}\Big\} \leq nP\Big\{|\sum_i Z_{ij} - EZ_{ij}| \geq \epsilon nh a_{n\vartheta}\Big\}$$

then (60) would follow if we could show that for any x,

$$P\Big\{\sup_{\substack{\alpha\in B_n^{(1)}\\ \beta\in B_n^{(2)}}}|\sum_i R_{ix}(a,b)|\geq \epsilon nha_{n\vartheta}\Big\}=O(n^{-a})\text{ for some }a>2, \tag{61}$$

where $B_n^{(1)}=\{a\in R:|a-a_x|\leq ca_{n\vartheta}\}$, $B_n^{(2)}=\{b\in R:|b-b_x|\leq c\{a_{n\vartheta}+(nh/\log n)^{1/2}/h\}\}$, $a_x=m(\theta_0^\top x)$, $b_x=m'(\theta_0^\top x)$, $R_{ix}(a,b)=Z_{ix}(a,b)-EZ_{ix}(a,b)$, $K_{ix}=K(X_{ix}^\top\vartheta/h)$ and $Z_{ix}(a,b)=K_{ix}X_{ix}[\varphi(Y_i-a_x-b_x\theta_0^\top X_{ix})-\varphi(Y_i-a-b\theta_0^\top X_{ix})]$. To this end, partition $B_n^{(i)},\ i=1,2$ into a sequence of sub rectangles $D_1^{(i)},\cdots,D_{J_1}^{(i)},\ i=1,2$ such that

$$|D_{j_1}^{(i)}|=\sup\Big\{|a-a'|:a,a'\in D_{j_1}^{(i)}\Big\}\leq M_n^{(i)}/M,\quad 1\leq j_1\leq J_1,$$

where $M_n^{(1)}=ca_{n\vartheta}$, $M_n^{(2)}=c\{a_{n\vartheta}+(nh/\log n)^{1/2}/h\}, M\equiv\epsilon^{-1}$ and $J_1\leq M$. Choose a point $a_{j_1}\in D_{j_1}^{(1)}$ and $b_{k_1}\in D_{k_1}^{(2)}$. Then

$$\begin{aligned}\sup_{\substack{a\in B_n^{(1)}\\ b\in B_n^{(2)}}}|\sum_i R_{ix}(a,b)| &\leq \max_{1\leq j_1,k_1\leq J_1}\sup_{\substack{a\in D_{j_1}^{(1)},\\ b\in D_{k_1}^{(2)}}}|\sum_{i=1}^n\{R_{ix}(a_{j_1},b_{k_1})-R_{ix}(a,b)\}|\\ &\quad+\max_{1\leq j_1,k_1\leq J_1}|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|\equiv H_{n1}+H_{n2}.\end{aligned}\tag{62}$$

We first consider H_{n2}.

$$P\Big\{H_{n2}\geq\frac{\varepsilon nha_{n\vartheta}}{2}\Big\}\leq J_1^2P\Big\{|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|\geq\frac{\epsilon nha_{n\vartheta}}{2}\Big\}$$

As $R_{ix}(a_{j_1},b_{k_1})$ is bounded and $\mathrm{Var}\{R_{ix}(a_{j_1},b_{k_1})\}=O\{h(a_{n\vartheta}+(nh/\log n)^{-1/2}\}$, then by Bernstein's inequality we have

$$J_1^2P\Big\{|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|\geq\frac{\epsilon nha_{n\vartheta}}{2}\Big\}\leq CJ_1^2\exp\{-\epsilon^2n^{1/2}h^{3/2}\}=O(n^{-a}),$$

for some $a>2$.

We next consider H_{n1}. For each $j_1=1,\cdots,J_1$ and $i=1,2$, partition each rectangle $D_{j_1}^{(i)}$ further into a sequence of subrectangles $D_{j_1,1}^{(i)},\cdots,D_{j_1,J_2}^{(i)}$. Repeat this process recursively as follows. Suppose after the lth round, we get a sequence of rectangles $D_{j_1,j_2,\cdots,j_l}^{(i)}$ with $1\leq j_k\leq J_k,\ 1\leq k\leq l$, then in the $(l+1)$th round, each rectangle $D_{j_1,j_2,\cdots,j_l}^{(i)}$ is partitioned into a sequence of subrectangles $\{D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)},1\leq j_l\leq J_l\}$ such that

$$|D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}|=\sup\Big\{|a-a'|:a,a'\in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}\Big\}\leq M_n^{(i)}/M^{l+1},\ 1\leq j_{l+1}\leq J_{l+1},$$

where $J_{l+1} \leq M$. End this process after the (L_n+2)th round, with L_n being the smallest integer such that

$$(2/M)^{L_n} > a_{n\vartheta}/M_{n\vartheta}^{(2)} \text{ [which means } 2^{L_n} \leq \{M_{n\vartheta}^{(2)}/a_{n\vartheta}\}^{\log(M/2)/\log 2}]. \tag{63}$$

Let $D_l^{(i)}$, $i=1,2$, denote the set of all subrectangles of $D_0^{(i)}$ after the lth round of partition and a typical element $D_{j_1,j_2,\cdots,j_l}^{(i)}$ of $D_l^{(i)}$ is denoted as $D_{(j_l)}^{(i)}$. Choose a point $a_{(j_l)} \in D_{(j_l)}^{(1)}$ and $b_{(j_l)} \in D_{(j_l)}^{(2)}$ and define

$$V_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\Big|\sum_{i=1}^n \{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a_{j_{l+1}},b_{k_{l+1}})\}\Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{l+1}}\Big\},\ 1 \leq l \leq \mathrm{L}_n+1,$$

$$Q_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\sup_{\substack{a\,\in\, D_{(j_l)}^{(1)},\\ b\,\in\, D_{(k_l)}^{(2)}}} \Big|\sum_{i=1}^n \{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b)\}\Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{l}}\Big\},\ 1 \leq l \leq \mathrm{L}_n+2.$$

Then $Q_l \leq V_l + Q_{l+1},\ 1 \leq l \leq L_n+1$. We first give a bound for $V_l,\ 1 \leq l \leq L_n+1$. As $R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a_{j_{l+1}},b_{k_{l+1}})$ is bounded and

$$E|R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a_{j_{l+1}},b_{k_{l+1}})|^2 \leq h\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}/M^{l+1},$$

applying Bernstein's inequality and using (63), we have

$$V_l \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big)\exp[-\epsilon^2 nh \min\{a_{n\vartheta}, a_{n\vartheta}^2(nh/\log n)^{1/2}\}] \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big)\exp(-\epsilon^2 n^{1/2}h^{3/2}). \tag{64}$$

We now focus on Q_{L_n+2}. Recall the definition of $Z_{ix}(a,b)$

$$Z_{ix}(a,b) = K_{ix}[\varphi(Y_i - a_x - b_x\theta_0^\top X_{ix}) - \varphi(Y_i - a - b\theta_0^\top X_{ix})]X_{ix}.$$

For any $a \in D_{(j_l)}^{(1)}$ and $b \in D_{(k_l)}^{(2)}$, let $I_i^{a,b} = 1$, if there is a discontinuity point of $\varphi(.)$ between $Y_i - a_{j_l} - b_{k_l}\theta_0^\top X_{ix}$ and $Y_i - a - b\theta_0^\top X_{ix}$ and $I_i^{a,b} = 0$ otherwise. Write

$$R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b) = \{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b)\}I_i^{a,b} + \{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b)\}(1-I_i^{a,b}).$$

Then we have $|\{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b)\}(1-I_i^{a,b})| \leq C\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}/M^l$ and specifically for $l = \mathrm{L}_n+2$

$$P\Big\{\sup_{\substack{a\,\in\, D_{(j_l)}^{(1)},\\ b\,\in\, D_{(k_l)}^{(2)}}} \Big|\sum_{i=1}^n \{R_{ix}(a_{j_l},b_{k_l}) - R_{ix}(a,b)\}(1-I_i^{a,b})\Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{\mathrm{L}_n+3}}\Big\}$$

$$\leq P\Big\{\sum_{i=1}^n U_i \geq \frac{1}{8}Mnh\Big\} \leq P\Big\{\sum_{i=1}^n U_i - EU_i \geq \frac{Mnh}{16}\Big\}$$

where $U_i = I\{|X_{ix}^\top \vartheta| \le h\}$ and the first inequality is due to (63). By Bernstein's inequality, this in turn implies that for $l = L_n + 2$

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big) P\Big\{ \sup_{\substack{a \,\in\, D^{(1)}_{(j_l)},\\ b \,\in\, D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}(1 - I_i^{a,b}) \Big| \ge \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\} = O(n^{-a}), \qquad (65)$$

for some $a > 2$. Now we have to show similar result for

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big) P\Big\{ \sup_{\substack{a \,\in\, D^{(1)}_{(j_l)},\\ b \,\in\, D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \ge \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\},\ l = L_n + 2.$$

Note that for any $a \in D^{(1)}_{(j_l)}$ and $b \in D^{(2)}_{(k_l)}$, $I_i^{a,b} \le I\{Y_i \in S_i\}$, where

$$S_i = [a_{j_l} + b_{k_l}\theta_0^\top X_{ix} - CM_n^{(2)}/M^l, a_{j_l} + b_{k_l}\theta_0^\top X_{ix} + CM_n^{(2)}/M^l],$$

which is independent of a, b. Let $U_i = I\{|X_{ix}^\top \vartheta| \le h\} I\{Y_i \in S_i\}$. As $R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)$ is bounded, we have for $l = L_n + 2$,

$$\begin{aligned} &P\Big\{ \sup_{\substack{a \,\in\, D^{(1)}_{(j_l)},\\ b \,\in\, D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \ge \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\} \\ &\le P\Big\{ \sum_{i=1}^n U_i \ge \frac{\epsilon nh a_{n\vartheta}}{C 2^{\mathrm{L}_n+2}} \Big\} \le P\Big\{ \sum_{i=1}^n U_i - EU_i \ge \frac{\epsilon nh a_{n\vartheta}}{C 2^{\mathrm{L}_n+4}} \Big\}, \end{aligned} \qquad (66)$$

where the second inequality is due to (63). Applying Bernstein's inequality to the right hand side of (66) and by (63), we have

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big) P\Big\{ \sup_{\substack{a \,\in\, D^{(1)}_{(j_l)},\\ b \,\in\, D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \ge \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\} = O(n^{-a}),\ \text{for } l = L_n + 2$$

for some $a > 2$. This together with (65) implies that $Q_{L_n+2} = O(n^{-a})$ for some $a > 2$. Therefore, based on (70), we have

$$P\Big\{ H_{n2} \ge \frac{\epsilon nh a_{n\vartheta}}{2} \Big\} \le Q_1 \le \sum_{l=1}^{L_n+1} V_l + Q_{L_n+2} = O(n^{-a}),$$

for some $a > 2$. ■

Lemma 6.15 *For all large enough $M > 0$, $Q_2 \le M d_n$ a.s., where*

$$d_n = nh a_{n\vartheta}^2 l_n / h\{1 + a_{n\vartheta}^{-1}(nh/\log n)^{-1/2}\} = o(nh a_{n\vartheta}^2),$$

Proof Let $X_{ik} = X_i - x_k$, $\mu_{ik} = (1, X_{ik}^\top)^\top$, $K_{ik} = K(X_{ik}^\top \vartheta/h)$ and write $\Phi_{ni}(x_k; \alpha, \beta) - \Phi_{ni}(x; \alpha, \beta) = \xi_{i1} + \xi_{i2} + \xi_{i3}$, where

$$\xi_{i1} = \Big(K_{ik}\mu_{ik} - K_{ix}\mu_{ix}\Big)^\top \alpha \int_0^1 \{\varphi_{ni}(x_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(x_k; 0)\}\, dt,$$
$$\xi_{i2} = K_{ix}\mu_{ix}^\top \alpha \int_0^1 \{\varphi_{ni}(x_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(x; \mu_{ix}^\top(\beta + \alpha t))\}\, dt,$$
$$\xi_{i3} = K_{ix}\mu_{ix}^\top \alpha \{\varphi_{ni}(x; 0) - \varphi_{ni}(x_k; 0)\}.$$

Then $P(Q_2 > M^{3/2} d_n/3) \le \mathrm{T}_n(P_{n1} + P_{n2} + P_{n3})$, where

$$P_{nj} \equiv \max_{1 \le k \le \mathrm{T}_n} P\Big(\sup_{x \in \mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} |\sum_{i=1}^n \xi_{ij}| \ge M^{3/2} d_n/9\Big),\ j = 1, 2, 3.$$

Based on Borel-Cantelli lemma, $Q_2 \le M^{3/2} d_n$ almost surely, if $\sum_n \mathrm{T}_n P_{nj} < \infty$, $j = 1, 2, 3$. Again this can be accomplished through similar approach in Lemma 5.1 in Kong et al(2008). We only deal with P_{nj} to illustrate.

First note that if $\xi_{i1} \neq 0$,then either $K_{ik} \neq 0$ or $K_{ix} \neq 0$. Without loss of generality, suppose $K_{ik} \neq 0$, i.e. $|X_{ix}^\top \vartheta| \le h$, whence $|X_{ix}^\top \theta_0| \le h + |\delta_\vartheta|$ and $|\mu_{ik}^\top(\beta + \alpha t)| \le C\{M_{n\vartheta}^{(1)} + M_{n\vartheta}^{(2)}\}$.

For any fixed $\alpha \in B_n^{(1)}$ and $\beta \in B_n^{(2)}$, let $I_{ik}^{\alpha,\beta} = 1$. If there exists some $t \in [0, 1]$, such that there are discontinuity points of $\varphi(Y_i - a)$ between $\mu_{ik}^\top(\beta(x_k) + \beta + \alpha t))$ and $\mu_{ik}^\top \beta_p(x_k)$; and $I_{ik}^{\alpha,\beta} = 0$, otherwise. Write $\xi_{i1} = \xi_{i1} I_{ik}^{\alpha,\beta} + \xi_{i1}(1 - I_{ik}^{\alpha,\beta})$. As $|(K_{ik}\mu_{ik} - K_{ix}\mu_{ix})^\top \alpha| \le C M_{n\vartheta}^{(1)} l_n/h$ and $|\mu_{ik}^\top(\beta + \alpha t)| \le C M_{n\vartheta}^{(2)}$, we have

$$|\xi_{i1}(1 - I_{ik}^{\alpha,\beta})| \le C M_{n\vartheta}^1 M_{n\vartheta}^2 l_n/h = o(a_{n\vartheta}^2)$$

uniformly in i, α, β and $x \in \mathcal{D}_k$, if $nh^3/\log n^3 \to \infty$. Let $U_{ik} = I\{|X_{ik}^\top \vartheta| \le 2h\}$. As $\xi_{i1} = \xi_{i1} U_{ik}$ (because $l_n = o(h)$), we have

$$\begin{aligned} P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \sup_{x \in \mathcal{D}_k} \Big| \sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta}) \Big| > \frac{M d_n}{18}\Big) &\le P\Big(\sum_{i=1}^n U_{ik} > \frac{Mnh}{18C}\Big) \\ &\le P\Big(|\sum_{i=1}^n U_{ik} - EU_{ik}| > \frac{Mnh}{36C}\Big), \end{aligned} \tag{67}$$

where the second inequality follows from the fact that $EU_{ik} = O(h)$. We can then apply to (67) Bernstein's inequality for independent data or Lemma 5.4 in Kong et al (2008) for dependent case, to obtain the below result

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta}) \Big| > M d_n/18\Big) \text{ is summable over } n, \tag{68}$$

whence $\sum_n \mathrm{T}_n P_{n1} < \infty$, is equivalent to

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \xi_{i1} I_{ik}^{\alpha,\beta} \Big| > M d_n/18 \Big) \text{ is summable over } n. \tag{69}$$

To this end, first note that $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in S_{i;k}^{\alpha,\beta}\}$, where

$$\begin{aligned} S_{i;k}^{\alpha,\beta} &= \bigcup_{j=1}^m \bigcup_{t \in [0,1]} [a_j - A(X_i, x_k) + \mu_{ik}^\top(\beta + \alpha t), a_j - A(X_i, x_k)] \\ &\subseteq \bigcup_{j=1}^m [a_j - CM_{n\vartheta}^{(2)}, a_j + CM_{n\vartheta}^{(2)}] \equiv D_n, \quad \text{for some } C > 0, \\ A(x_1, x_2) &= m(x_1^\top \theta_0) - m(x_2^\top \theta_0) - m'(x_1^\top \theta_0)(x_1 - x_2)^\top \theta_0, \end{aligned}$$

where in the derivation of $S_{i;k}^{\alpha,\beta} \subseteq D_n$, we have used the fact that $|X_{ik}| \le 2h$, $\mu_{ik}^\top(\beta + \alpha t) = O(M_n^{(2)})$ and $A(X_i, x_k) = O(h^2 + |\delta_\vartheta|^2) = o(M_n^{(2)})$ uniformly in i. As $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in D_n\}$, we have $|\xi_{i1}| I_{ik}^{\alpha,\beta} \le |\xi_{i1}| U_{ni}$, where $U_{ni} \equiv I(|X_{ik}| \le 2h) I\{\varepsilon_i \in D_n\}$, which is independent of the choice of α and β. Therefore,

$$\begin{aligned} P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \xi_{i1} I_{ik}^{\alpha,\beta} \Big| > M d_n/18 \Big) &\le P\Big(\sum_{i=1}^n U_{ni} > MnhM_n^{(2)}/(18C) \Big) \\ &\le P\Big(\sum_{i=1}^n (U_{ni} - EU_{ni}) > \frac{MnhM_n^{(2)}}{36C} \Big), \end{aligned} \tag{70}$$

where the first inequality is because $|\xi_{i1}| \le CMa_{n\vartheta} l_n/h$ and the second one because $EU_{ni} = O(hM_n^{(4)})$. Similar to (67), we could apply either Bernstein's inequality for independent data or in dependent case Lemma 5.4 in Kong et al (2008) to see that (69) indeed holds. ∎

Lemma 6.16 *All eigenvalues of* $(S_2 + \theta_0\theta_0^\top)^{-1}(\Omega_0 + \theta_0\theta_0^\top)$ *fall into the interval* $(0, 1)$.

Proof By the Cauchy-Schwarz Inequality that for any $x \in R^d$,

$$\begin{aligned} &E\{g(X)(X - x)|X^\top\vartheta = x^\top\vartheta\} E\{g(X)(X - x)|X^\top\vartheta = x^\top\vartheta\}^\top \\ &\le E\{g(X)|X^\top\vartheta = x^\top\vartheta\} E\{g(X)(X - x)(X - x)^\top|X^\top\vartheta = x^\top\vartheta\}, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\{\nu_\vartheta(x) - x\mu_\vartheta(x)\}\{\nu_\vartheta(x) - x\mu_\vartheta(x)\}^\top \le \mu_\vartheta(x)\omega_\vartheta(x) \\ \text{or} \quad &\mu_\vartheta(x)\{(\nu/\mu)_\vartheta(x) - x\}\{(\nu/\mu)_\vartheta(x) - x\}^\top \le \omega_\vartheta(x). \end{aligned}$$

Multiply both sides by $m'(x_0^\vartheta)^2$ and take expectation, we have that $S_2 - \Omega_0 \geq 0$, which could be strengthen as $S_2 - \Omega_0 > 0$. This is because if there exists some $\vartheta_1 \neq 0$, such that $\vartheta_1^\top(S_2 - \Omega_0)\vartheta_1 = 0$, then for any x, there exists some C, such that

$$\begin{aligned}
&\{g(X)\}^{1/2}\vartheta_1^\top(X-x) \equiv C\{g(X)\}^{1/2}, \text{ for all } X^\top\vartheta = x^\top\vartheta \Rightarrow \\
&\vartheta_1^\top(X-x) \equiv C, \text{ for all } X^\top\vartheta = x^\top\vartheta \Rightarrow \vartheta_1 \equiv \vartheta \qquad (71)
\end{aligned}$$

A sufficient condition for $(S_2 + \theta_0\theta_0^\top)^{-1}(\Omega_0 + \theta_0\theta_0^\top)$ to have only positive eigenvalues is that θ_0 is the sole eigenvector of S_2 and Ω_0 that corresponds to eigenvalue 0. We argue this by contradiction. Suppose there exists some ϑ such that $\vartheta \perp \theta_0$ and

$$E\{g(X)\vartheta^\top(X-x)(X-x)^\top\vartheta|\theta_0^\top X = \theta_0^\top x\} = 0, \text{ for any } x \in R^d \qquad (72)$$

$$E\{g(X)\vartheta^\top(X-x)|\theta_0^\top X = \theta_0^\top x\} = 0, \text{ for any } x \in R^d \qquad (73)$$

Note that as $g(X) > 0$, (72) in fact implies that $E\{\vartheta^\top(X-x)|\theta_0^\top X = \theta_0^\top x\} = 0$, which in turn means that $\vartheta = \theta_0$; this contradicts the fact that $\vartheta \perp \theta_0$.

To show that (73) can't be true, let $\{b_1, \cdots, b_{d-1}\}$ constitute the orthogonal basis of the orthogonal space to vector θ_0. Let $x = b_i,\ i = 1, \cdots, d-1$, then $\theta_0^\top x = 0$ and from (73) we have

$$E\{g(X)\vartheta^\top(X-b_i)|\theta_0^\top X = 0\} = 0, \Rightarrow \vartheta^\top E\{g(X)X|\theta_0^\top X = 0\} = \vartheta^\top b_i E\{g(X)|\theta_0^\top X = 0\}$$

As $E\{g(X)X|\theta_0^\top X = 0\}$ and $E\{g(X)|\theta_0^\top X = 0\}$ are constants (vector) independent of b_i and $E\{g(X)X|\theta_0^\top X = 0\} \perp \theta_0$, we have that there exists some vector $b \perp \theta_0$ such that

$$\vartheta^\top b = \vartheta^\top b_i, \quad i = 1, \cdots, d-1, \Leftrightarrow \vartheta^\top(b - b_i) = 0 \quad i = 1, \cdots, d-1,$$

but this can not be true unless $\vartheta \perp b_i$ for all $i = 1, \cdots, d-1$.

Next we show that $\lambda_{max} < 1$ by contradiction. If not, suppose x is the corresponding eigenvector,

$$\begin{aligned}
&(S_2 + \theta_0\theta_0^\top)^{-1}(\Omega_0 + \theta_0\theta_0^\top)x = \lambda_{max}x \Rightarrow (\Omega_0 + \theta_0\theta_0^\top)x = \lambda_{max}(S_2 + \theta_0\theta_0^\top)x \\
&\Rightarrow x^\top(\Omega_0 + \theta_0\theta_0^\top)x = \lambda_{max}x^\top(S_2 + \theta_0\theta_0^\top)x \Rightarrow x^\top\Omega_0 x \geq \lambda_{max}x^\top S_2 x (\because \lambda_{max}x \geq 1)
\end{aligned}$$

which contradicts the fact that $S_2 - \Omega_0 > 0$ if $x \neq \theta_0$.

REFERENCES

Andrews, D.W.K. (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62**, 43-72.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer, New York.

Chaudhuri, P., Doksum, K. and Samarov, A. (1997) On average derivative quantile regression. *Ann. Statist.* **25**, 715-44.

Koenker, R. (2005) *Quantile Regression.* Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978) Regression Quantiles. Econometrica. **46** (1), 3350.

Kong, E., Linton, O. and Xia, Y. (2008) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model, *Econometric Theory* (to appear)

Korolyuk, V. S. and Borovskikh, Yu. V. (1989) Law of the iterated logarithm for U-statistics. *Ukrainian Mathematical Journal* **1**, 89-92.

Lange, T. Rahbek, A. and Jensen, S.T. (2006) Estimation and Asymptotic Inference in the First Order AR-ARCH Model. Preprint

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186-99.

Sun, S. and Chiang, C.Y. (1997) Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. Journal. Applied Mathematics and Stochastic Analysis. **10**, 3-20.

Welsh, A.H. (1996) Robust estimation of smooth regression and spread functions and their derivatives *Statistica Sinica* **6**, 347-366.

Xia, Y. (2007) Direct estimation of the multiple-index model